

Networks for Africa

The future of science and technology in Africa depends on the development of mutually supportive networks. Two examples show how imaginative initiatives can be turned into models for others.

South Africa occupies a curious political position within its continent. The wealthiest of all the African nations, it inspires outsiders with the manner of its transition over the past 11 years from an apartheid-based regime towards a 'rainbow nation'. The vision of the African National Congress, as promoted so successfully by Nelson Mandela, is now being carried forward by his successors. But because of its past or its ability to prosper from the wealth and infrastructure built up by the apartheid regime, South Africa is sometimes viewed a little askance by other African nations, and its capacity to take a lead may be compromised.

But that is no reason for its initiatives to receive anything less than encouragement from other African countries, especially when they have the potential to be extended beyond its borders. And given that most donor support is directed at health and agriculture, indigenous initiatives that strengthen mathematical and physical sciences in the region are to be especially welcomed. Two in particular are worth celebrating and deserve further support.

Star performers

One example is the National Astrophysics and Space Science Programme. Its main aim is to make coherent use of a small and fragmented academic community distributed around South Africa's universities, to provide education and training. The scheme is based at the University of Cape Town and coordinated there, but all the universities involved take responsibility for developing and delivering courses and hosting students for research projects. Lecturers from other universities travel to Cape Town to give course modules. The other universities do not lose out, however — more than 50% of MSc students return to the partner institutions to complete their course, and any university can award the degree.

The approach was developed by the astronomy community itself and attracted support from the Ford and Andrew W. Mellon foundations. It is now rightly seen by the government as a model to be encouraged, and there is hope that South Africa's Department of Education will provide funds to support it and other similar schemes.

Importantly, the scheme has attracted students from other African countries. As accounts of its success feed back, so more students from those countries apply. One might hope that in any of these countries, several years from now, trained researchers returning from the scheme could have accumulated enough critical mass to set up centres in their own countries. The students are selected for their potential to do research — this is African capacity building in action. Other disciplines and other countries take note.

In a suburb of Cape Town is another example, also kick-started with the help of philanthropy: the African Institute for Mathematical Sciences (AIMS). The institute was launched in 2003 in what had been a dilapidated hotel, donated by the leading ANC member Ben Turok and his cosmologist son Neil, who now chairs its governing

body. Neil Turok's original proposal met much scepticism, but theoretical physicist Fritz Hahne, former dean of science at Stellenbosch University, saw the potential and became its director.

Fully residential, it accommodates some 40 graduate students from across Africa and is overflowing with academics from the developed world willing to give their time to come and teach them. There is an emphasis on applying mathematics to practical problem solving and learning to use back-of-the-envelope thinking to develop both analytical and computational approaches. Topics studied in more depth range from immunological and metabolic modelling to control theory, stochastic calculus and fundamental physics and cosmology. Vocational skills are also taught to introduce communication techniques for students, whose native languages include French and Arabic.

"Although students initially have little in common, a pan-African feeling quickly develops."

Most of the students on this year's course have moved on to South African universities to take postgraduate degrees, with about ten returning home. Although AIMS students initially have little in common, given the diversity of their countries and backgrounds, a pan-African feeling quickly develops, according to Hahne, and contacts with alumni have stayed strong.

AIMS costs about US\$800,000 a year, and hopes to expand to take in 70 students. It has drawn on philanthropic funds from the Mellon and Gatsby foundations and from Vodafone, and also on the commitment of those who give their time to work there — staff, students and visitors, including young African researchers.

Future aims

The operation's success has led to a proposal to set up 15 institutes across Africa over the next five years — not clones of AIMS but suitably tailored by those willing to take the lead in each case. This African Mathematical Institutes Network would hope to deliver 300 trained postgraduate mathematical scientists every year. The cost of establishing such a network could come to about \$20 million over the period.

This proposal depends for its implementation on the science and technology segment of the New Partnership for Africa's Development (NEPAD), set up in 2001 as an agency by which African states can channel development activities. NEPAD is set to be the key agency by which the G8 industrialized nations intend to fulfil their ambitions to increase assistance to Africa. NEPAD's "consolidated plan of action" for science and technology (www.nepadst.org/publications/docs/doc27_082005.pdf) includes several flagship projects in biotechnology, water, information technology and materials. A network fostering generically useful skills in mathematical sciences fully deserves to sit alongside them. ■



The heat is on

A successor to the Kyoto Protocol on climate change must involve mandatory emissions caps.

Talks about a climate accord to succeed the Kyoto Protocol when it expires in 2012 begin in earnest next week in Montreal. They will take place amid concerns that nations who backed the protocol are retreating from its central principle: the imposition of mandatory caps on greenhouse-gas emissions.

No national leader still in office is more strongly associated with the Kyoto agreement than Britain's prime minister, Tony Blair, and his recent pronouncements on Kyoto II have worried supporters of mandatory caps. In a series of speeches earlier this month, Blair made no mention of targets and echoed US President George W. Bush by stressing the role of technology development in cutting emissions. Blair also said that something "better and more sensitive" than the initial agreement was needed to convince major developing nations such as India and China, which do not have to limit emissions under the current protocol, to sign up to a new version.

That seems fair enough. But if European leaders such as Blair fail to insist on targets as part of Kyoto II, there is a danger that the entire exercise could become meaningless. Technology, in the shape of cleaner fossil-fuel power stations, renewable energy sources and perhaps nuclear power, ought to form an important element of nations' climate-change strategies. But these technologies need to be nurtured through financial incentives produced by mandatory caps and carbon-trading arrangements.

Following criticism of his initial remarks, Blair has been talking up targets again, stating that "targets, sensitively and intelligently applied over the right timeframe" are needed after 2012. But it will take remarkable ingenuity to bridge the chasm between developed countries, such as the United States and Australia, that have done

little to cut their own emissions, and developing ones, such as China and India, that want rich nations to act before they do.

That said, there are already ideas in circulation about how to bring on board all these parties, including the United States, where a new administration elected in 2008 may take a more constructive approach. Specific industrial sectors might, for example, be asked to accept targets. China might agree to set targets on its energy-intensive cement industry, which has substantial greenhouse-gas emissions, in return for more technical support from overseas companies, who would earn credits that they could trade off against the emissions commitments at home. Similar schemes already operate on a small scale under the Kyoto Protocol.

Other sectors could be regulated on an international basis. Governments might agree to establish national targets on vehicle fuel emissions and efficiency, for example. This would offer nations the chance to sign up to agreements in sectors in which they know they can improve without losing their competitive advantage.

Working out how these ideas can be combined with the existing carbon-trading system will be an immense challenge. The last thing the process needs is for nations already committed to emissions targets under the original Kyoto protocol to turn their backs on them now. It was relatively painless for some nations — notably Britain and Germany — to meet tight Kyoto targets, because local events had sharply reduced emissions shortly after 1990, the baseline date against which the protocol's targets were set.

Now new circumstances, including greater electricity demand in southern Europe and steady economic growth, are making it harder for the European Union to stay within the Kyoto caps. Its leaders must redouble their efforts to restrict emissions and to vigorously pursue as strong a successor agreement as is practicable. ■

"The last thing the process needs is for nations already committed to emissions targets under Kyoto to turn their backs on them now."

Life is what you make it

This issue celebrates the emerging field of synthetic biology.

How's this for creativity? Take *Escherichia coli* bacteria. Transform them into light-sensitive organisms by fusing a photoreceptor from the cyanobacterium *Synechocystis* to a protein in the *E. coli* membrane. Make a film (in both senses) of such bacteria and use them to record an image with a resolution of 100 megapixels per square inch. For the result, see page 441. For other bio-widgits, see page 417.

Achieving this neat trick required researchers to engineer component parts of gene circuitry. This bottom-up engineering is often referred to as synthetic biology. It is indeed a type of biology: developing circuits that achieve what nature has evolved over eons is one way of gaining insight into what makes life tick. But it is also engineering, of a type quite different to the simple manipulation of bits of DNA to incorporate or knock out existing genes. This technology

allows biological components, circuits and potentially replicating organisms to be developed from scratch, possibly based on different genetic codes from those found in the wild.

In this special issue on synthetic biology, some of the field's founding figures describe the technical challenges of such engineering (see page 443), review the scope and foundational principles of the discipline (page 449), and explore ways in which socially responsible synthetic biologists can gain public trust by focusing on safety (page 423).

Last year we expressed the hope that synthetic biologists would act to engage with stakeholders (*Nature* 431, 613; 2004). *Nature* is pleased to highlight community thinking on such issues, and welcomes, and will participate in, stakeholder discussions at the second international meeting on synthetic biology at Berkeley next May. And finally, knowing that a graphic can help get the message across to a wider audience, we are delighted to publish a cartoon introduction to the field (www.nature.com/nature/comics/syntheticbiologycomic). ■

"This technology allows biological components, circuits and potentially replicating organisms to be developed from scratch."

RESEARCH HIGHLIGHTS

CELL BIOLOGY

Slug beats puma

Cell **123**, 641–653 (2005)

Damage to a cell's DNA activates the tumour suppressor protein p53, which can either initiate DNA repair or trigger cell suicide. A protein called Slug has been found to ensure that p53 acts as a saviour in some tissues by preventing the action of *puma*, a gene that promotes cell death.

After finding this link *in vitro*, researchers led by Thomas Look of the Dana-Farber Cancer Institute in Boston, Massachusetts, showed that mouse mutants unable to make Slug died of bone-marrow failure after exposure to doses of radiation that are not lethal to normal mice. The team suggests that finding a way to boost Slug levels could permit stronger doses of chemo- and radiotherapies to be used without damaging bone marrow.

MATERIALS

An aluminium lotus

J. Am. Chem. Soc. **127**, 15670–15671 (2005)

A lotus leaf shakes off water because its bumpy, waxy surface makes it super-hydrophobic: water droplets just roll off. Mimicking this effect in synthetic micro- and nanostructured coatings has become commonplace, and could lead to mist-proof windows, self-cleaning surfaces and solid lubricants. But such coverings have tended to be exotic and expensive.

Weimin Liu of the Lanzhou Institute of Chemical Physics in China and his co-workers have recreated the 'lotus effect' cheaply on aluminium and its alloys. They simply spin-coat the metals with a fluorinated hydrocarbon or a hydrophobic polymer and heat them in a vacuum. This produces finely textured surface films that water cannot wet.

PHYSIOLOGY

Light clockwork

Cell Metabol. **2**, 297–307 (2005)

A study has illuminated some of the pathways through which light resets the body clock and controls the hormonal system.

Researchers headed by Hitoshi Okamura of Kobe University, Japan, investigated how exposure to light influences the activity of the adrenal glands. These organs sit above the kidneys in humans and secrete hormones that are involved in regulating metabolism.

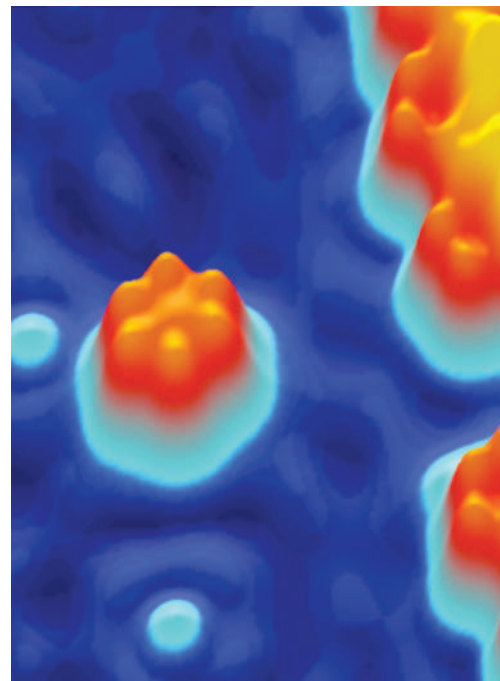
The team's experiments, carried out in mice, reveal that a part of the brain called the suprachiasmatic nucleus directly controls

Tools of the trade

Nature Mater. doi:10.1038/nmat1529 (2005)

Researchers in Europe have designed a dumper truck for nanoscale construction: it's a molecule that can capture, carry and unload up to six atoms at a time.

The molecule consists of a central benzene ring from which six bulky legs protrude (pictured). The legs hold up the ring, creating a cage underneath. Francesca Moresco of the Free University of Berlin, Germany, and her colleagues used the tip of a scanning tunnelling microscope (STM) to drive the molecule over lone atoms on a copper surface. The legs guided the atoms into the central cage where they became trapped until the molecule was lifted from the surface. The team says this approach could be used to assemble nanostructures on insulating surfaces, which are difficult to manipulate using standard STM techniques.



F. MORESCO ET AL.

the activity of the adrenal glands, a process previously thought to be mediated by the brain's pituitary gland. The researchers also found that exposure to light increased expression of the *Per1* gene — believed to be a key player in the circuitry of the biological clock — in the adrenal glands.

GENOMICS

Coffee's cousin

Theor. Appl. Genet. doi:10.1007/s00122-005-0112-2 (2005)

Two of the world's best-loved crops are more similar than geneticists had previously realized. A new database of

genes from the coffee plant *Coffea canephora* shows that it has an almost perfect gene-for-gene match with the tomato plant *Solanum lycopersicum*.

Of the more than 13,000 genes that make up the coffee genome, almost all have counterparts that perform the same function in tomato, as well as in other solanaceous species, such as potato and aubergine. Researchers led by Steven Tanksley of Cornell University in Ithaca, New York, made the discovery after analysing the genetic sequences expressed in coffee seeds at a range of developmental stages, and comparing them with previously published sequences for other species.

QUANTUM PHYSICS

The third man

Phys. Rev. Lett. **95**, 200502 (2005)

Encoding information in entangled quantum states — for example, by polarizing individual photons — allows tamper-proof communication between two parties. But this offers more security than some would like. In the cryptography schemes demonstrated by Yu-Ao Chen and his colleagues, a boss wants to retain control over the ability of his two employees to exchange secure information.

The researchers, at the Hefei National Laboratory for Physical Sciences at the Microscale in China, use quantum optics to carry out two different protocols that enable

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

CUBOIMAGES SRL/ALAMY

this. In both cases, the employees can establish a secret key for decoding encrypted data only with their boss's cooperation. These three-party protocols are made possible by the team's mastery of four-photon entanglement.

PALAEOBIOLOGY

Dung grasses up dinosaurs

Science **310**, 1177–1180 (2005)

Grasses have come to dominate habitats across the world since the dinosaurs' rule. Yet, because the earliest unequivocal grass fossils came from some 10 million years after the end of the dinosaurs' era, the two were never thought to have met.

Now we have proof that they did. Researchers led by Caroline Strömberg from the Swedish Museum of Natural History in Stockholm have discovered silica structures characteristic of grass in fossilized dinosaur droppings from 65 million years ago. The dung is suspected to have come from titanosaur sauropods (pictured right). Although grass was not a major part of the dinosaurs' diet, the team suggests it may have been important for early mammals that had teeth similar to today's grazers.

CELL BIOLOGY

In miniature

Nature Struct. Mol. Biol. doi:10.1038/nsmb1019 (2005)

The enzyme telomerase plays an important role in cells, stitching short stretches of DNA on to the ends of chromosomes to stabilize them. But study of the telomerase from yeast, a model organism, has been hindered because the enzyme contains a large RNA component that misfolds in the test tube.

Thomas Cech and his colleagues at the University of Colorado in Boulder have now put together a miniature version of the RNA unit, Mini-T, that makes reconstitution of the active enzyme possible. Mini-T functions well in yeast cells, although such cells are less fit than those with the normal enzyme.

This system will allow cell biologists to characterize the yeast telomerase mechanism by adding back bits of the RNA unit, in addition to Mini-T, to see what happens.

MICROFLUIDICS

Air-breathing for energy

J. Am. Chem. Soc. doi:10.1021/ja054599k (2005)

US researchers have developed an air-breathing microfluidic fuel cell that could prove cheaper than conventional alternatives, while rivaling their performance.

Microfluidic fuel cells, because they handle small volumes of fluid, can use the principle of laminar flow to separate the fuel and its oxidant, whereas conventional cells must use a physical barrier — a costly membrane — to control the interaction of the two fluids.

But researchers exploring this approach found that the cell's power output was limited by the rate that dissolved oxygen, the oxidant, could diffuse through the cell. Led by Larry Markoski of INI Power Systems in Cary, North Carolina, and Paul Kenis from the University of Illinois at Urbana-Champaign, the team now reports that using an electrode that is permeable to atmospheric oxygen overcomes the problem.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

TECTONICS

Plate spinning

Geology **33**, 857–860 (2005)

A novel explanation for the spinning of tectonic plate fragments — seen where one plate sinks beneath another — has emerged from real-time measurements of the motion.

Laura Wallace of the Institute of Geological and Nuclear Science in New Zealand and her colleagues used a global positioning system to characterize the rotation of microplates, with millimetre accuracy, at five subduction zones. These rotations occur in the upper plate, they conclude, when a buoyant mass such as an island or undersea ridge blocks the downward flow of the lower plate. The collision creates the fragments, then the pieces are set spinning by the torque that the lower plate exerts as its unblocked regions continue to subduct. The researchers suggest that this mechanism may also trigger the stretching of the upper plate that sometimes causes it to split — a phenomenon known as backarc rifting.

JOURNAL CLUB

Ronald Breaker

Yale University, New Haven, Connecticut

A biochemist speculates on the battle between proteins and RNA to control the cell.

The 'RNA world' theory about life's origin argues that RNA alone ran the metabolism of the first cells. As a supporter of this theory, I like to think that some of the rare RNA molecules, such as ribozymes, that persist in cells today are direct descendants of ancient RNA organisms.

If so, these molecules have survived billions of years of stiff evolutionary competition from proteins, which dominate the machinery of modern cells. This is testament to the power of RNA.

But I have to confront the reality of evolution. If RNA is so good, why have protein enzymes driven ribozymes, their 'RNA world' counterparts, to the brink of extinction?

RNA molecules are known to have deficiencies: they have limited chemical complexity, for example, and are notorious for spontaneously self-destructing. Also, Watson and Crick taught us that each nucleotide has a complementary partner, which means that ribozymes at high concentrations could experience antisense catastrophes as molecules bind to others with matching sequences.

Researchers at the Whitehead Institute in Cambridge, Massachusetts, recently added another deficiency to the list. They showed that existing ribozymes have a difficult time evolving new activity (E. Curtis & D. Bartel *Nature Struct. Mol. Biol.* **12**, 994–1000; 2005). Proteins, however, can yield new catalysts with only a few mutations. The fitness landscape in which ribozymes evolve seems to be much more jagged.

I still believe that RNA is a functionally powerful polymer, but these new findings lead me to suspect that ancient organisms had a bumpy evolutionary ride until proteins emerged to pave the road ahead.

A. HAYWARD/ARDEA.COM

NEWS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

AP/P. SAKUMA

Google makes data free for all

Google launched a new service last week, Google Base. It allows anyone to upload files for free to its massive server farms, making the data instantly searchable. Although mainly aimed at online markets for such things as homes and jobs, scientists say the facility could have important implications for data-sharing in science, and perhaps boost efforts to make the web more 'intelligent'.

As well as letting people upload data, Google Base lets users describe the data with simple tags that others can then use in searches. It also allows users to structure the data by adding fields on the fly. So a web page holding a scientific article might have fields for 'author', 'journal', 'publication date' and other bibliometric information.

That might not sound like a very big deal. But advocates say that this allows web content to be structured as databases on a large scale. For a start, that makes it simple for any scientist to share data, and store it in ways that allow computers to search and retrieve it.

Scientists are still "in the Dark Ages" when it comes to sharing data, says David Haussler, director of the Center for Biomolecular Science and Engineering at the University of California at Santa Cruz. Data falling outside the relatively few big international databases, such as those for gene sequences, protein structures and astronomy data, mainly end up in supplementary tables accompanying journal articles, he says, and are "stored in some non-indexable, inconsistent and inconvenient format, if indeed they are kept at all".

"To have a way to easily cross-examine multiple kinds of data would be a real boon to research."

Google Base or similar services could help, says Ian Foster, a computer scientist at Argonne National Laboratory in Illinois and co-inventor of the Grid concept, in which many com-

puters work together to provide large amounts of processing power and data storage. Science badly needs "something that would make it trivial for individuals and communities to create and share scientific data, and the programs that operate on those data", he says.

"To have a way to easily cross-examine multiple kinds and sources of data would be a real boon to research," agrees Paul Myers, a bioinformatician from the University of Minnesota, Morris. "I think Google is getting in early on what could be an immensely important tool."

Smart systems

Google Base may also signal a modest start for the web to move towards the 'intelligent' network originally envisaged by Tim Berners-Lee when he invented the web at CERN, the European Laboratory for Particle Physics in Geneva, Switzerland, in 1989.

Most web pages are designed to be read by humans, and don't contain additional descriptive information that can be interpreted by computers. This limits their usefulness, especially for users carrying out searches. For example, it's not currently possible to search the web to find "only peer-reviewed papers dealing with experiments where the CCR5 protein activates the PYK2 protein". And when reading a paper online, you can't ask the computer to replot a graph adding in extra data sets.



GENETICISTS DESIGN MADCAP MICE
Study of risky rodents could improve drugs for anxiety
www.nature.com/news

Berners-Lee champions what he calls a 'semantic web', where tags added to pages would allow computers to 'understand' what the pages contain. This means computers can ask whether the data meet certain criteria and merge data sets from different sources.

But although the semantic web is fast gaining ground in certain specialist areas such as bioinformatics, it has yet to take off in a big way. Scientists say Google Base could change that by bringing structured web pages to the masses. "The big issue here is whether services like this will help bootstrap the semantic web," says Greg Tyrelle, a proteomics researcher at Chang Guan University in Taiwan.

Google power

"Flexible online storage of arbitrary data, including scientific data, is going to be a major area of research over the next couple of years," says Leigh Dodds, a web expert at publisher Ingenta. "Google Base takes that a step further by widening it out to everyone," although he adds that he would like to see governments and universities doing more to promote such services, rather than leaving it to Google.

Scientists point out, however, that Google has been prominent in its absence from work on the semantic web in the World Wide Web consortium (W3C), the body that creates web standards. They also acknowledge that Google Base is a pretty crude service so far, especially compared with sophisticated specialist databases such as GenBank and UniProt. All you can do is put in information, and then search it — there's no way to extract or compute the data.

But most researchers believe that will change fast. Google has been a pioneer in creating what are known as 'application programming interfaces' to its other services, such as Google Maps. These allow anyone to write programs that can access Google's databases, and mix and match its content with other data to create completely new products.

"If Google wants to turn Google Base into more than just a tool for finding information, and into something scientists can actually use to explore data, then more is needed," says Mark Gernstein, a bioinformatician at Yale University in New Haven, Connecticut.

But observers such as Foster believe such progress could happen fast. "Google has much relevant technology and expertise," he says. "If it forms the right partnerships and dedicates sufficient resources, it could have a tremendous impact."

"Google Base looks a little simple right now, and it's not clear exactly how to tap into Google's power," adds Myers. "But we've got to start somewhere."

Declan Butler

US watchdog finds bias against morning-after pill

WASHINGTON DC

A US government watchdog has announced that there was a high-level effort within the US Food and Drug Administration (FDA) to overrule agency scientists and block over-the-counter access to an emergency contraceptive.

Many critics suspected as much when over-the-counter access to Plan B (levonorgestrel) was denied in 2004 despite the recommendations of agency scientists and outside experts. But a government report issued on 14 November has now reached the same conclusion.

The Government Accountability Office (GAO), the investigative arm of Congress, says the effort was led by the then commissioner of the FDA, Mark McClellan, and note that McClellan's involvement along with his high-level colleagues in what is normally a staff decision was "unusual".

Arthur Caplan, director of the Center for Bioethics at the University of Pennsylvania, told *Nature* that in 20 years of watching and advising the FDA, "I've never seen this level of high-up administrative intervention into the scientific review process. The science indisputably supports approval, but the politics didn't. I think the higher-ups including the then commissioner put the politics first."

Plan B works by preventing fertilization or implantation if it is taken within 72 hours of intercourse. Conservatives adamantly opposed making the contraceptive accessible without a prescription, fearing it would increase

teenage promiscuity.

In 2003 an FDA advisory committee of external experts voted 23–4 to make Plan B available over-the-counter, a change backed by staff reviewers at the agency. But the GAO reports that McClellan repeatedly resisted approval of the move, because of concerns about its use by younger teenagers, despite data showing that the safety issues were no

made a decision nor did he recommend a decision" on Plan B.

Susan Wood, a former assistant FDA commissioner who quit in September in protest over the FDA's handling of Plan B, said in a statement: "This report is a sad reminder of why I felt compelled to resign — that the FDA leadership is ignoring its process and not relying on science and medical evidence."

The Bush administration has been under increasing fire for politicizing scientific decisions on matters from global warming to the reliability of condoms. Last week, attention also focused on upcoming recommendations for vaccination against cervical cancer. A vaccine has been shown to be effective in preventing the cancer, which is caused by the sexually transmitted human papilloma virus. Conservative groups argue that widespread vaccination will encourage sexual promiscuity among young people.

In an effort to pre-empt these groups' influence on the upcoming vaccination policy, Senator Hillary Clinton (Democrat, New York), who also spoke out against the FDA's treatment of Plan B, wrote last week to Mike Leavitt, the Health and Human Services secretary. The Bush administration, Clinton said in the letter, "has repeatedly allowed ideology, not science, to form the basis of policy... We do not want to see another instance of ideology trumping the health and well-being of the American people."

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Former FDA chief Mark McClellan denies blocking access to Plan B contraceptive.

different for younger women.

In May 2004, two months after McClellan had been replaced by acting commissioner Lester Crawford, the change was rejected. The manufacturer of Plan B, Barr Laboratories, was encouraged to resubmit an application to make Plan B available to women over 16. But this August the agency indefinitely postponed the decision on that application (see *Nature* 437, 179; 2005).

The FDA responded to the GAO report saying it "mischaracterizes facts... We question the integrity of the investigative process." Gary Karr, spokesman for McClellan also insists that the former FDA chief "neither

ON THE RECORD

“There’s a perception that we don’t bring much to the party.”

John LaMattina, of drug company Pfizer, laments the fact that only 9% of the US public believe the pharmaceutical industry is honest.

“She’s not under any pressure, she goes at her own steady pace... and is loved by everybody.”

A spokesman for Australia Zoo explains the longevity of Harriet the tortoise, believed to be the world’s oldest animal, who just turned 175.

Sources: *New York Times*, *BBC News*

SCORECARD

 **Minke whales**
Scientists have finally identified minke whales as the source of a mysterious ‘boing’ sound, first heard some 50 years ago in the North Pacific Ocean.

 **Reality television**
British TV producers have created elaborate sets for a show — including speakers to play rocket noises, and decor from the film *Space Cowboys* — to con participants that they have been launched into space.

 **Trust**
Support for President George W. Bush among US scientists and engineers has sunk to just 6%, down from 30% four years ago.

NUMBER CRUNCH

Bat guano offers some surprising nutritional benefits compared with a Big Mac. Or at least, that’s one possible reading of a study of the dietary habits of cave salamanders, which like to feast on the former.

54% of guano is protein.

23% of a Big Mac is protein.

1% of guano is fat.

33% of a Big Mac is fat.

Source: Fenolio, D. B. et al. *Proc. R. Soc. Lond. B* doi:10.1098/rspb.2005.3341 (2005).

Neuroscientists put gene therapy into reverse

WASHINGTON DC

Gene therapy has attracted plenty of fanfare but provided very little in terms of positive results. Giving people new genes to remedy defects in their old ones turns out to be a difficult business. But solid if little-noticed progress is being made in an approach that turns the concept on its head: rather than curing conditions, researchers are finding ways to study brain disease by inserting faulty genes into healthy animals.

Several pioneers of the technique presented their latest results at this year’s annual meeting of the Society for Neuroscience, held on 12–16 November in Washington DC. They say the animal models they have created are easier to make than those based on alternative methods, mimic human versions of disease better,

and can be applied to a wider range of species. “They open up possibilities that other models don’t allow,” says Anders Bjorklund, a specialist in neural transplantation at the University of Lund in Sweden.

Some of the most well-developed models are for Huntington’s disease, a fatal movement disorder caused by a single faulty gene. The mutant gene expresses altered huntingtin protein, which results in damaged brain cells. One of the first Huntington animal models to be created by gene therapy was achieved in 2002 by Nicole Déglon of the Atomic Energy Commission in Orsay, France (*J. Neurosci.* **22**, 3473–3483; 2002). Déglon injected a virus containing the faulty huntingtin gene into the striatum of rats. The virus inserted itself into neural DNA, causing expression of the huntingtin protein

Seals net data from cold seas

Marine mammals are ready to start collecting data on the remote polar seas, say biologists. They have been tagging animals as part of their work for decades, and now they say the satellite tags are sophisticated enough to provide oceanographers with valuable data too.

Mike Fedak of the Sea Mammal Research Unit (SMRU) says he hopes the tags will bring together oceanographers and marine biologists. “One of the most exciting things is that we have something to talk about with oceanographers, and they actually want to hear it,” he says.

The standard probes used by physical oceanographers are called CTDs, because they record conductivity, temperature and density. When dropped on lines or towed behind ships, they allow researchers to calculate the salinity and depth of the surrounding water, to track the distribution and movement

of water masses.

The problem is that researchers are largely limited to areas where ships already travel. When it comes to the remote polar regions, especially the Antarctic, there is a dearth of data.

“There are a few floats out there but not much else,” says Lars Bohme, an oceanographer who has been working with Fedak’s team at the University of St Andrews, UK. “There’s a need for real-time data from this area, to develop models that in the short term can help predict weather and in the long term, climate change.”

Fedak and his colleagues worked with engineers at the SMRU to develop satellite tags containing miniaturized CTDs that can be carried by whales and seals. After a preliminary test on two beluga whales (*Delphinapterus leucas*) in Norway, a larger study called SEaOS (Southern elephant seals as Oceanographic Samplers)

has convinced the researchers that the data are good enough to meet oceanographers’ exacting standards.

The fist-sized tags were attached to the heads of about 70 elephant seals (*Mirounga leonina*) at four breeding areas around the Southern Ocean over the past three years. During the Antarctic winter, the seals dive roughly 40 times a day to between 300 and 800 metres, to hunt for fish and squid. When the animal surfaces, data collected from its dive are transmitted to the Argos system of satellites. These relay the information to researchers in real time.

“The tags provide precise, accurate data,” says Bohme. He is using information from SEaOS to profile the Southern Ocean’s currents, and says he is now popular with other oceanographers, who are keen for data from the under-sampled Southern Ocean.

“Oceanographers treated it

and tissue damage typical of the disease.

Compared with other techniques for creating transgenic animal models, such as adding or deleting genes from embryonic cells and then breeding the animals that develop, the process is quick. "We can create these models in months rather than years," says Déglon. "And we can replicate the damage seen in late stages of disease, which is not often seen in other transgenic models."

Déglon is now writing up results from a macaque Huntington model. She says her team has observed motor deficits such as muscle contractions that are typical of the condition, and that post-mortem analysis of brain tissue shows the gene causes damage similar to that seen in humans.

The ability to create primate models using this technique has persuaded others to use viral vectors. Deniz Kirik, a colleague of Bjorklund's at Lund, is finishing a three-year study in which he tracked marmosets infected with a mutant version of the gene for alpha-synuclein — the protein involved in Parkinson's disease. He says the model, the first transgenic primate model

of the disease, recreates human symptoms well. The technique also allowed him to use a new control in his experiment: by injecting the faulty gene into just one side of the marmosets' brain, he was able to show that movement problems only developed on the side of the body controlled by that part of the brain.

Other primate models of Parkinson's have used chemicals to kill the neurons that are normally damaged by the disease. "This model is much better — the pathology is driven by mechanisms similar to those in humans," says Kirik.

Such advantages mean that the use of viral-vector models is likely to grow beyond the ten or so groups that have taken them up over the past few years. But Bjorklund notes that the technique will not replace traditional transgenic models. For example, conventional techniques allow large numbers of animals with identical mutations to be bred, which is useful for drug screening. The surgical skills needed to inject the virus into the correct brain areas are also difficult, adds Bjorklund, and may put off some labs. ■

Jim Giles



M. BIJOU

Ahead of the pack: seals with satellite tags could gather data from oceans where few ships venture.

with scepticism at first because it seemed wacky," adds Fedak. "But tests of the tag's accuracy have shown this can work."

Jamie Morison, an oceanographer at the University of Washington's Polar Science Center in Seattle

is convinced. "This is a really useful idea," he says, although he adds that seals will not be able to do everything. "I'm interested in processes underneath the ice," he says. "These animals won't go where the ice is too thick for them to

come up and breathe."

SEaOS will finish collecting data in January 2006. Fedak hopes to work with researchers in eight other countries to use seal and whale species in hard-to-reach polar areas. ■

Tom Simonite



THE NATURE PODCAST
Listen to the stories behind
the hottest research in our
weekly show, available free
at [www.nature.com/
nature/podcast](http://www.nature.com/nature/podcast)

Korean stem-cell crisis deepens

The pressure on stem-cell pioneer Woo Suk Hwang over the way he obtained human eggs for his research is intensifying — particularly in South Korea, where he had been a national hero. In the past week, several new claims have emerged that Hwang may have used eggs that were paid for, as well as eggs from junior members of his laboratory.

Hwang's team, based at Seoul National University, has produced a string of landmark papers in stem-cell research, including the first stem cells obtained from a cloned human embryo (W. S. Hwang *et al. Science* **303**, 1669–1674; 2004) and the first patient-matched embryonic stem cells (W. S. Hwang *et al. Science* **308**, 1777–1783; 2005).

Recently, his research has been overshadowed by allegations about the way he obtained eggs, but until now these have come from outside his home country. In 2004, *Nature* published a claim that Hwang's group had used eggs from one of his graduate students — a charge Hwang has constantly denied (see *Nature* **429**, 3; 2004). The student later withdrew her claim. Then two weeks ago, Hwang's close friend and collaborator Gerald Schatten of the University of Pittsburgh, Pennsylvania, broke off ties. He accused Hwang of possible ethical irregularities and misrepresentations regarding egg donation, although he gave no other details (see *Nature* **438**, 262–263; 2005).

Now further claims of possible impropriety are arising closer to home. On 21 November, Sun Il Roh, a fertility expert at MizMedi Hospital in Seoul, gave a press conference at which he admitted that 20 eggs that he had procured and given to Hwang for his 2004 study were paid for. According to Korean newspaper reports, Roh said he paid 1.5 million won (US\$1,430) of his own money to each of the 20 women whose eggs were used in the experiment. Quoted in the *JoongAng Daily*, he says: "This is not a large amount of money, considering that they had to receive injections every day for 8–10 days." But Roh, who was a co-author on Hwang's 2005 paper, insists that Hwang did not know the status of the eggs he received. Roh did not respond to *Nature's* requests for an interview.

Although buying eggs for research was not illegal when the eggs were procured in 2003, the practice is hugely controversial, and has been illegal in Korea since last January. Supplementary material to Hwang's 2004 paper clearly states that all egg donors were volunteers.

On 22 November, as *Nature* went to press, Seoul-based Munhwa Broadcasting Corporation (MBC) was to run an investigative

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Fertility specialist Sun Il Roh admits at a press conference to having paid women to donate eggs.

programme that it said would show further evidence that Hwang used eggs from junior members of his lab. Obtaining eggs from a graduate student is problematic because of the risk that personal pressure could be applied.

According to an MBC producer, the programme was to produce medical records of egg donors from MizMedi Hospital. The records allegedly show that at least one of the donors was a researcher in Hwang's lab. MBC says the researcher in question is not the student who last year told *Nature* she had donated eggs at MizMedi, before withdrawing the claim.

Furthermore, an informer, who MBC says is closely linked to Hwang's lab, allegedly provided the station with experimental notes containing the donor's name, patient number and the date the eggs were used. MBC claims these details match the MizMedi medical reports.

The *Chosun Ilbo* on Tuesday cited another source claiming that eggs from two researchers in Hwang's lab were used. One was "a graduate student who invented a new way of removing the nucleus from eggs and is now working at a research institute in a US university", it said.

Hwang has not replied to *Nature's* repeated requests for an interview.

The effects of the allegations on the stem-cell field and on Hwang's research are unclear. On 15 November, after the news of Schatten's separation, the Korean government laid out plans to invest 11.5 billion won in the World Stem Cell

Hub, an international research network to have been led by Hwang. But it also proposed to detach the hub from Seoul National University and make it an independent body. Many potential overseas collaborators have said their plans are on hold until the allegations are resolved.

How these events will affect Hwang's team's ability to publish is another open topic. "It's pretty clear that the editor of any journal would be on heightened alert if they received a piece of work from them, and would probably scrutinize it very carefully for the ethics of the work at the very least," says Gregory Curfman, executive editor of *The New England Journal of Medicine*.

"If it turned out that deliberate falsifications had been communicated to us in connection with that paper we would certainly have to make an announcement of that," says Donald Kennedy, editor-in-chief of *Science*, which published the work. "We would certainly say something about the caution with which we would treat future communications from that group."

Most are holding out for a revelatory statement from Schatten or Hwang, who has been rumoured to be planning a press conference this week. "I hope Dr Hwang will give us the whole story," says Kennedy. "There's a different onus on Schatten — he has issued a dramatic statement and he's leaving it like a dead seal on our collective desks."

David Cyranoski

Additional reporting by Erika Check

AFP/GETTY IMAGES

Deal on toxicity law fails to appease

MUNICH

An uneasy compromise has been struck to see Europe's chemical-safety law off the starting blocks. The European Parliament last week gave a preliminary go-ahead to legislation that will bring in extensive toxicity testing of both new and existing chemicals.

But the deal struck during the vote, which would see fewer chemicals tested than originally planned and a drive towards animal-free tests, has left most stakeholders dissatisfied.

The controversial law, known as REACH (Registration, Evaluation and Authorization of Chemicals), is one of the most complicated pieces of legislation that Europe has attempted (see *Nature* 438, 144–146; 2005). It raises the standards for safety testing of chemicals and puts full responsibility for safety assurance in the hands of industry.

Chemicals produced at levels of more than 1,000 tonnes per year will have to go through a full range of tests that for each compound will cost about €2 million (US\$2.3 million) and will use some 4,000 animals.

REACH replaces 1981 legislation that required data on the toxicity of new chemicals to be registered. But its range will also extend to some 30,000 chemicals that were in commercial use before 1981, for which no safety data are currently listed. Most of these compounds will now have to be tested.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Activists argue that Europe's toxicity tests will kill too many animals.

Since being proposed by the European Commission more than two years ago, REACH has been the focus of intense lobbying, particularly by industry, which argues that it will be too expensive, and by animal-rights supporters, who complain that the testing will cost too many animal lives.

The compromises introduced at the parliamentary reading on 17 November aim to meet some of these objections. Less safety information than planned will now be demanded for compounds in chemical classes that are generally agreed to be non-toxic and are produced in relatively low amounts each year. Of these,

any produced at less than 10 tonnes per year will be exempt from all testing.

In response to animal-welfare concerns, the revised law increases the influence of the European Centre for the Validation of Alternative Methods (ECVAM) in approving testing strategies that do not use animals. For example, it gives ECVAM, based in Ispra, Italy, the right to check that a manufacturer is not using animal tests when alternatives exist. And it requires that regulators adopt any alternative test validated by ECVAM within 14 days — something that currently takes months or years.

REACH also obliges industry to share data on safety testing so that each compound is tested only once. A new European Chemicals Agency will be established in Helsinki, Finland, to administer the legislation.

But animal-welfare groups complain that the changes don't go far enough in reducing the number of animals needed for testing. And industry groups maintain that the tests and administration remain prohibitively expensive. Environmental groups, meanwhile, say that the compromises mean that some potentially dangerous chemicals will go untested.

The parliamentary vote is the first step in a process during which the draft law can still be altered. REACH is scheduled for approval next year and should be implemented in 2007. ■

Alison Abbott

P. GOETGHELUCK/SPL

China steps up drive to vaccinate all domestic birds

China claims to be well on the way towards vaccinating every domestic bird in the country against avian flu. The bold scheme — which would mean inoculating some 14 billion birds — was announced on 15 November and comes in response to outbreaks that animal-health officials say are dangerously widespread. On 16 November, China confirmed its first two human cases of bird flu, one of which was fatal.

The agricultural ministry says it began a large-scale compulsory vaccination programme in 2004, covering outbreak areas and places considered to be at high risk. According to a ministry official,

some 8 billion birds (60% of China's domestic bird population) have already been vaccinated. The use of vaccines in specific areas has successfully reduced the number of outbreaks, says Fusheng Guo, the avian-flu surveillance network coordinator of the Food and Agriculture Organization (FAO) in China.

But Guo says that the number of outbreaks, which reached 17 on 21 November, means that China is approaching an "emergency situation". And that makes mass vaccination even more urgent. "It's a good idea if you can do it properly, with surveillance

afterwards," Guo says.

Can China complete such a huge programme? Yes, according to Guo, who says that the country's ten vaccine producers can make 16 billion doses of vaccine per year. If need be, he says, they could double that amount by doubling workers' shifts.

The FAO's senior officer for the Infectious Disease Group in Rome, Juan Lubroth, says that China's technology and research in vaccines is top notch — including use of reverse genetics, in which a section of the virus's gene responsible for virulence is removed.

But Lubroth warns that care must be taken to maintain the quality of

the vaccines. "If the vaccine is substandard, you won't get any protection where you think you did." The 'vaccine brigades' must also wash clothing and equipment so they don't spread the virus, he says.

There has been scepticism over whether it will be possible to deliver the vaccines to all of the birds kept in small backyard farms. "The logistics of herding in all those loose chickens and ducks is a little more difficult," says Lubroth. But China has proved itself with its belated but impressive response to SARS, says Guo: "Policemen and soldiers can help in some areas." ■

David Cyranoski

**BIRD FLU**

Visit our up-to-date blog to get all the latest developments on avian influenza

www.nature.com/news

Software shakes up schizophrenia diagnosis

Computer analysis of brain images can diagnose schizophrenia in patients, possibly even before symptoms arise, say researchers at the University of Pennsylvania in Philadelphia.

The technique, which is based on the ability of computers to tease out subtle differences between brain images, has split neuroscientists, with some questioning the value of the information produced.

But Ruben Gur and his colleagues are convinced it works. Earlier this year, they claimed that they could use the technique to detect whether individuals are lying or telling the truth (see *Nature* **437**, 457; 2005). Now they have turned their attention to mental disorders.

They used magnetic resonance imaging (MRI) to scan the brains of 69 schizophrenia patients and 79 healthy controls. The images were analysed by computer to produce an algorithm that could tell the two groups apart. Rather than focusing on specific areas of the

brain thought to be affected by the disorder, as has been tried in the past, they looked for subtle changes across the whole brain.

This type of approach has proved successful before — but only for images used to derive the algorithm. As soon as fresh images were introduced, the success rate plummeted. But this time the researchers say that they have overcome this problem and that they were able to classify new individuals as schizophrenic or healthy with 81% accuracy (C. Davatzikos *et al. Arch. Gen. Psychiatry* **62**, 1218–1227; 2005).

Gur makes bold claims for the technique, saying that it is ready to be used alongside clinical histories and interviews to help diagnose schizophrenia. And, because many of the people studied were 18 years old and in the early stages of the disease, he believes that it might be possible to use brain imaging to diagnose the disease before classic symptoms appear.

“Now we can give a computer a picture of a

person's brain and ask whether or not this person has schizophrenia,” says Gur. “This should do for schizophrenia what the echocardiogram did for heart disease.”

Other experts are much more cautious. “The results are interesting and promising, but need replication,” says Philip McGuire, an expert in brain scanning and schizophrenia at the Institute of Psychiatry in London.

“Finding differences compared with healthy people is the easy part,” says Chris Frith, who specializes in schizophrenia and imaging at University College London. What will be useful, he says, is a way to distinguish people with schizophrenia from those suffering from related problems such as mania or severe depression.

Gur and his team now plan to compare schizophrenics with other mentally ill people. They also want to compare young sufferers with their family members.

Jennifer Wild

Experts say US authorities should change patent laws

US universities and government agencies have much work to do to ensure that large-scale patenting of genes and proteins doesn't impede biomedical research, an expert panel says.

The increasingly complex maze of patents, licences and material-transfer agreements could slow research considerably, a committee of lawyers, scientists and business experts reported on 17 November. For instance, universities might restrict their investigators' activities in the wake of a 2002 court ruling, which held that researchers are not necessarily protected from patent-infringement charges — even if they do not intend to profit from their work.

The experts, organized by the National Academy of Sciences, recommended that patent standards be tightened and undergo scientific review, that researchers should avoid seeking patents on non-therapeutic genes and proteins, and that universities should streamline material-transfer agreements and protect their rights to distribute research materials. They also said Congress should consider passing laws to exempt universities and researchers from charges of patent infringement while doing basic research.

Europe presses ahead with plan to grab the headlines

The European Commission last week launched a programme to help European science news compete more strongly with news from the United States. The Communiqué initiative will develop a system to improve underperforming press operations in European research institutions and give European scientists the support they need to hit the headlines.

Launched in Brussels, the service is likely to involve a new central press association and initiatives to encourage European press officers to approach researchers more actively. Its organizers, which include the AlphaGalileo Foundation, Europe's Internet press centre for research in science and the arts, hope to obtain funding under the next four-year European Union Framework programme, starting in 2007.

Japanese craft hits trouble during asteroid landing

Japan's Hayabusa spacecraft failed to land on the asteroid Itokawa as planned this week, raising concerns about whether it will ever bring back dust samples.

On 12 November, the craft lost contact

Grizzlies set to lose protected status

Grizzly bears may soon lose some federal protection around one of their most famous locations, Yellowstone National Park.

On 15 November, the US Fish and Wildlife Service proposed removing the Yellowstone population of grizzlies (*Ursus arctos*) from the list of threatened and endangered species. Since 1975, when it was first listed, Yellowstone's grizzly population has grown from 220 bears to 600. More than 50,000 of the animals used to roam the United States. The proposal would not affect bears in the national park itself, but would affect those living around its boundaries in Wyoming, Montana and Idaho.

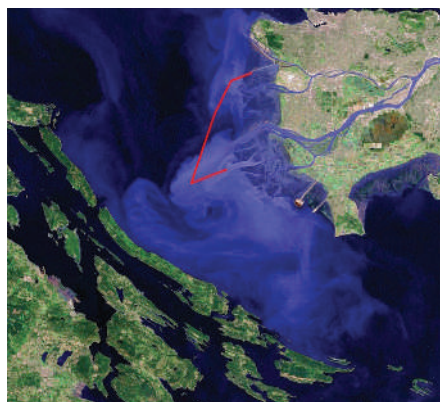
As the bears leave the list, another species is set to join. Last week, the government proposed listing a population of about 90 killer whales (*Orcinus orca*) that live in Washington state's Puget Sound. But Congress is currently considering legislation that would lessen the

with a tiny robot probe called Minerva. And on 20 November, Hayabusa experienced problems when it was less than 17 metres from its target. Scientists at the Japan Aerospace Exploration Agency (JAXA) say it seemed to slide sideways along the asteroid, at a height of about 10 metres, for more than 30 minutes. The agency also temporarily lost contact with the craft. Mission scientists sent a signal for the craft to ascend, and Hayabusa's position-adjusting sensor made the craft skyrocket to as far as 100 kilometres from Itokawa.

As *Nature* went to press, Japan's space agency was considering whether to attempt a second landing.

Plans to monitor undersea activity start to take shape

Storms, earthquakes and even schools of fish will be monitored by an undersea observatory system near Vancouver, Canada, that is due to go online next month. Known as the Victoria Experimental Network Under the Sea (VENUS), this real-time system will use an array of sensors and



Laying the cable: the VENUS system will use monitors laid under the sea near Vancouver.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

protections conferred by the Endangered Species Act. A final decision on the whales and bears will be made in mid-February, after a period for public comment.

video cameras. A website will display live data and images.

The first phase of the Can\$10.3-million (US\$8.5-million) project will be four kilometres of cable laid next month under Patricia Bay. A 40-kilometre cable will then be positioned in the Strait of Georgia, one of Canada's busiest waterways.

Study assesses virtues of a home-grown agency boss

A survey has quantified the notion that government leaders are more effective if they climb the ranks of an agency, rather than being appointed to the top job from outside. David Lewis, a political scientist at Princeton University in New Jersey, used a management-rating tool to assess whether a programme had clear goals and strategies, and whether it was achieving its objectives.

The survey has a 100-point scale, and the average ranking is 61.7. Programmes headed by political appointees rate five or six points lower than those headed by career-based employees, Lewis reports.

Lewis also used the system to examine research and development (R&D) programmes in US agencies for *Nature*. "The effect is a little stronger," he says. Of the R&D agencies, the National Science Foundation programme dealing with logistics for polar research had the highest score, 95.3. The Small Business Innovation Research/Technology Transfer programme at the Department of Defense had the lowest score, just 23.5.

Correction

Our News story on the rift between stem-cell researchers Gerald Schatten and Woo Suk Hwang (see *Nature* 438, 262–263; 2005) misstated which partner gave an effusive public toast to the other. Schatten praised Hwang, not vice versa.

POWER STRUGGLE

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

For decades, California has bucked the US trend of gobbling ever more electricity. But can the state pull off an even more ambitious goal and slash its greenhouse-gas emissions? **Charles Petit** finds out.

Arnold Schwarzenegger has a mission: he wants to terminate global warming. In June, the California governor called for the state to slash its greenhouse-gas emissions to 80% of 1990 levels in the next 45 years. "The debate is over," he said in a forthright speech in San Francisco. "We know the science. We see the threat. And we know the time for action is now."

This was fighting talk, but if any advanced economy can pull off such drastic cuts in emissions, this high-technology Pacific Rim state and its 36 million residents probably can. Schwarzenegger has help. His muscle for the job is a team of state energy-conservation experts who have been in the business for years. And first among them is his polar opposite: a short, skinny physicist named Arthur Rosenfeld. More than three decades ago, Rosenfeld helped to trigger the state's successful fight to cut energy consumption; today he is one of the five members of California's Energy Commission.

Rosenfeld was Enrico Fermi's last graduate

student and he spent decades as a physics professor at the University of California, Berkeley. He now commutes weekly between his home overlooking San Francisco Bay and Sacramento, the capital, in an energy-saving Prius hybrid sedan that the state provides. The Energy Commission's job isn't easy: to help the most populous US state figure out how it might cut greenhouse-gas emissions and make money doing it.

Under control

In his office, Rosenfeld pulls out a data plot of which he is particularly fond. It shows electricity consumption per capita from 1960 to 2002, with one line for California and one for the United States. In 1960, both lines sit at 4,000 kilowatt-hours per person. They rise at roughly the same pace to about 7,000 kilowatt-hours in the early 1970s. But at the point when the US energy crisis struck that decade, the lines diverge dramatically: California virtually flatlines its energy use per citizen — even though its economy was outpacing the rest of

the nation. The state's electricity use per capita today is the lowest in the nation at 6,800 kilowatt-hours, compared with 12,800 kilowatt-hours for the country overall.

The strategies that helped California achieve those conservation goals may now help it in its greenhouse-gas cuts. State energy experts, including Rosenfeld, don't foresee California adopting many radical new technologies to meet its ambitious goals. Rather, a steady application of proven technologies should do much of the job.

California's \$1.5-trillion gross annual product makes it the world's sixth largest economy, behind France and ahead of Italy. It is the planet's ninth-largest emitter of greenhouse gases. "California is not an insignificant actor, and we are seen as a world leader in protecting the environment," says Eileen Tutt, a senior officer at the California Environmental Protection Agency. "Not to take action sends a very strong signal."

Still, the governor's pledge, made on the United Nations World Environment Day,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Harnessing the power of renewable energy such as wind (above) and geothermal (below), California aims to cut its greenhouse-gas emissions.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

caused more than a few jaws to drop. Schwarzenegger is a tax-cutting Republican who owns several fuel-slurping Hummers and is deeply suspicious of government regulation. Beset by budget fights and union opposition, he has sagged in popularity with the state's generally Democratic voters since his election two years ago. But his energy policies, building on those of a string of governors of both parties, get him kudos from longtime activists. "The governor is a real-life climate action hero today," Nancy Ryan, a senior economist with the group Environmental Defense, told reporters after his speech.

Specifically, Schwarzenegger vowed that California will cut its greenhouse-gas emissions to below 2000 levels by 2010 and to less than the 1990 level of 373 million tonnes by 2020. But then the governor added the final, ambitious goal to cut emissions by a further 80% by 2050.

Out on a limb

His policy stands in stark contrast to that of the federal administration under President George W. Bush, who has refused to ratify the Kyoto Protocol to reduce greenhouse-gas emissions. The president has said that such action would squeeze the US economy too much. California officials say that they can do it while boosting the economy and creating jobs. The state's strong environmental policies in the past, they point out, occurred while its economy thrived.

Success will require the cooperation of several interlocking agencies. The Energy Commission plays a major role, as do the state's Environmental Protection Agency, Air Resources Board and Public Utilities Commission. Schwarzenegger's proclamation renewed their "absolute licence to go out and make California a model country for greenhouse policies", says Stephen Schneider, a physicist and climate-policy analyst at Stanford University.

State officials have much at stake. California's climate could change utterly if a warmer world redirected storm paths (see *Nature* **430**, 818; 2004). Rising temperatures could cause winter rain instead of snow in the Sierra Nevada mountains, triggering floods for which the state's aqueducts and dams are not prepared. Plus, its coast is vulnerable to a rise in sea level.

Other states have also recognized their vulnerability to climate change, and have independently taken climate policy into their own hands. Local legislators, from mayors of cities to state governors, have begun their own versions of Kyoto-like regulations (see B. Fisher and R. Costanza *Nature* **438**, 301–302; 2005). In the northeast, nine states have agreed to cap carbon dioxide emissions from more than 600 power plants in the region (see *Nature* **437**, 11; 2005). On the west coast, California has joined

with Oregon and Washington in a governors' initiative to encourage energy efficiency and conservation.

But of all the states, California has set itself up as a model of how diligent attention to efficiency can take root and pay off. Its example has caught on: in recent years many other states have adopted California's standards for car pollution rather than the more lax federal standards.

And the state is now attracting international attention. In September, its Public Utilities Commission, Energy Commission and the Pacific Gas and Electric Company signed a pact with China's Jiangsu province to train officials and utility executives in energy-conservation tactics. Earlier this month, Schwarzenegger led a sales delegation to China to tout the state's energy-saving technologies, and another team from the state's Air Resources Board travelled to Belgium to brief European air-quality experts on energy policies.

California's approach to energy conservation has helped it save money. The state sets electricity rates for private utilities, and sometimes provides subsidies to help power companies induce customers to cut their consumption. If they do, the state gives money back to the companies — through rate adjustments and other payments — that makes up for what the firms would have earned had they built additional power plants.

The Energy Commission calculates that the total power bill for residents is about \$16 billion

lower each year than if the state had not launched its conservation campaign. Conservation has also managed to prevent some 18 million tonnes of carbon pollution being emitted from power plants — equivalent to taking 12 million cars off the roads. After allowing for the cost of measures such as changed building practices, appliances and subsidies, the net saving is about \$12 billion.

And deeper energy cuts should pay more, the commission says. The Air Resources Board estimates that planned reductions in greenhouse-gas emissions by 2020, from motor vehicles alone, could save Californians \$256 million annually by 2010 (mostly from smaller fuel bills), and \$4.8 billion annually by 2020.

Cut and dried?

But will the state's longer-term emissions policy succeed? Schneider is unsure how cost-effective the whole plan will be. Earlier stages may pay for themselves, he says, but the final leap to the 80% cut is unlikely to come without costs. "It would take a total revamp of our fuel infrastructure," he notes.

So far, even state planners aren't sure how they will meet the later goals. "We don't have the details, but we'll have a report to the governor's office in January," says Tutt.

Some fresh ideas are already in the works.

**"The governor is a
real-life climate
action hero today."
— Nancy Ryan**

One notion, set out by Schwarzenegger's administration, is to place 1 million solar-panel systems on rooftops by 2018. California gets about 11% of its electricity from geothermal, wind, biomass and solar units; for the United States overall, the number is around 2%. California aims to increase its share of renewable sources to 20% by 2010 and to 33% by 2020.

Also helpful will be the vehicle clean-up legislation enacted just before Schwarzenegger's arrival. This requires car manufacturers, starting in 2009, to cut greenhouse-gas emissions from new cars and trucks by 22% by 2013 and 33% by 2017. But the law remains in dispute — perhaps predictably, car companies have sued. They argue that carbon dioxide is not a pollutant, and that regulating it at state level would pre-empt federal control over the fuel-efficiency standards in new cars. In the long run, the governor has chosen hydrogen-fuelled cars as his personal crusade.

Wind power figures large in state plans. California pioneered wide-scale use of it and already has more than 14,000 wind turbines. In a good breeze their combined capacity is 2,100 megawatts — about the same as two nuclear power plants. State energy officials estimate that wind alone, in principle, can generate an additional 30,000 megawatts.

Fuel for thought

The rest of the renewable energy would probably come from beefed-up geothermal, solar and biomass facilities. In recent weeks, the utility company serving San Diego contracted to buy another 205.5 megawatts of wind power, and the Los Angeles utility company ordered 500 megawatts of solar power from a complex being built some 120 kilometres northeast of the city. When construction is completed in 2013, this will be the world's largest solar facility and will more than double US solar-energy generation. One energy source that is not on the agenda, despite not emitting greenhouse gases, is nuclear power, which now provides about 10% of the state's electricity. California law forbids the construction of additional nuclear plants until safe, long-term waste storage has been assured.

If Schwarzenegger's emissions-busting goal seems ambitious, Rosenfeld is happy to put it in perspective. He recalls the early days of conservation work when he co-founded a programme in energy-efficient buildings at the Lawrence Berkeley National Laboratory. The effort spawned what today is an entire Environmental Energy Technologies Division at the lab.

Back then, Rosenfeld's personal project was to make fluorescent lights more efficient. Around the same time, in 1973, state lawmakers formed the California Energy Commission, which was signed into existence by then-Governor Ronald Reagan.

In its early days, the commission took almost as gospel any reports from Rosenfeld's

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Arnold Schwarzenegger refuels a hydrogen-powered car (top). His vision is to see vehicles like this replace the polluting models on the road.

group on how to reduce energy demand. In 1978, the state used the team's recommendations as a guide to impose energy requirements on new buildings. Appliance standards came two years later governing such items as gas furnaces, air conditioning and refrigerators. Today, due largely to such regulations, refrigerators use a third of the electricity and cost about a third as much to buy as in the early 1970s. In contrast to California's efforts, federal standards, such as 'Energy Star' labels for approved appliances, did not arrive until the 1990s.

Energetic response

The total electricity used for air-conditioning new homes in California is now a third of what it was in the 1970s. And two technologies from the Lawrence Berkeley initiative — the revamped light bulb and window coatings that keep heat out in the summer and in during winter — will save the US economy some \$23 billion in utility bills, the National Academy of Sciences has estimated.

Rosenfeld is not done waging war on wasted energy. He is pushing hard for regulations to encourage the use of white roofs, or at least coloured ones, that reflect most near-infrared radiation. Such cool roofs could save \$200 million yearly in air-conditioning costs in Los Angeles alone. He frets over energy "vampires", the trickle of power that modern appliances continue to suck through their plugs even when they are turned off. Such vampires include televisions waiting for a signal from the remote, cable boxes, modems, recharging cell phones, cordless phones, garage-door openers waiting for a signal and stereos. Standby power, which was insignificant two decades ago, now typically accounts for 10% of a home's electricity use. New technology that gives appliances more efficient standby modes, which California will require as it becomes available, could reduce this by 75%.

"Let me show you something," Rosenfeld says as he pulls out an iconic NASA satellite image of the Western Hemisphere at night. The US industrial centres blaze like constellations against the black backdrop of the continent. "You are looking at millions of light bulbs," he says in dismay. To Rosenfeld, the necklaces of light are not signs of advanced civilization but of a wastrel society leaking precious photons into the void.

Then, Rosenfeld leans forward with a conspiratorial air. "In the next 20 years," he declares, "California will disappear." And that is Rosenfeld's dream — that the state will disappear, or at least fade, at night. Rules enacted this year require new lights for streets, parking areas and the like to focus 98% of their illumination on the ground, not into the sky. Once again, Rosenfeld says, the message for California on saving energy is simple: every little bit helps.

Charles Petit is a freelance writer in California.

S. YEATER/AP

N. UT/AP

Seeing in the dark

When darkness falls for Antarctica's long winter months, the sky becomes a spectacular canopy of stars. At one brand new base, astronomers are braving the extreme cold to build telescopes that they hope will rival space observatories.

Gabrielle Walker investigates.



K. AGABI

For the thirteen residents of Antarctica's newest scientific base, the winter is finally over. In early November, after nine months of isolation in temperatures dropping to 80 degrees below zero, the first relief planes arrived bearing fresh food, letters and — perhaps the biggest blessing of all — new faces.

On the relatively balmy Antarctic coastline, there are many permanent bases, but Concordia is the first year-round station to be built in the interior for 50 years. The site has been used by summer visitors since 1995 for a European ice-coring project, which was successfully completed last December¹. This year, the first winter crew spent most of the dark months working to finish the wiring and plumbing of their brand new home — a joint venture between the French Polar Institute Paul Emile Victor (IPEV) and the Italian Antarctic Research Programme (PNRA).

But they haven't laboured solely on construction jobs. Astronomer Karim Agabi from the University of Nice in France, part of Concordia's first winter crew, has good news to report about the site's potential for optical astronomy. Although more problematic than first thought, Concordia could be one of the best places in the world to site an optical telescope. The stars shine uncommonly steadily there, Agabi says. "You really have the impression that they're stuck to the sky".

The high Antarctic plateau has many potential advantages for astronomy, such as little atmospheric water vapour to distort incoming

light. The US National Science Foundation, based in Arlington, Virginia, has been exploiting this potential for decades at the South Pole, which proved to be excellent for observing microwaves, but astronomers have been frustrated in their attempts to observe ordinary visible light. One reason is the frequent auroras, which interrupt any serious star gazing.

Eternal night

A bigger problem is that the South Pole lies on a slope, so winds tumble down from on high, generating a layer of turbulent air more than 200 metres thick. Turbulent winds aggravate what astronomers call 'seeing', a measure of the jitter they observe in starlight as it passes through packets of air of different temperatures before reaching the telescope.

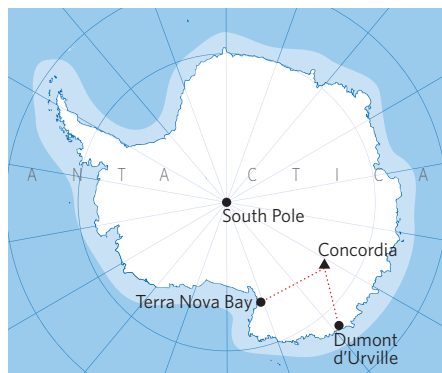
Concordia, in contrast, sits on a flat-topped dome of ice where there is almost no wind. The base is also outside the normal region for

auroras. And unlike other high spots in Antarctica, Concordia is relatively easy to reach. During the summer, people and sensitive equipment can be transported by ski-equipped planes from the Italian coastal station at Terra Nova Bay. Meanwhile, heavier material can be hauled over the ice by tractors from the French coastal base of Dumont d'Urville, 1,100 kilometres away (see map).

Many astronomers hoped that the seeing at Concordia would rival that available in space, but be much cheaper, with the additional benefit that humans would be on hand to fix equipment; transport costs on the traverse are just a few euros per kilogram, compared with many thousands for a space mission.

Agabi and other researchers from the University of Nice have been testing the site for the past five summers. Their white telescopes stand on a stylish 8-metre platform, graced with an arch modelled on the base of the Eiffel Tower — a design that one of the astronomers, Eric Aristidi, calls "the French touch". Alas, the architect's original vision was compromised a little when he persuaded the firms providing materials to hand over two of everything for the price of one, so now two golden arches stand side-by-side. As Aristidi acknowledges: "The idea was to copy the Eiffel Tower, not McDonald's".

The group's summer results seemed promising, using bright test stars visible despite the 24-hour daylight. Although seeing was poor at midnight, when the sun was at its lowest, by mid-afternoon it was significantly better than at any other ground-based site. There were hopes



LUAN



Reaching for the stars: the telescope platforms at Concordia were modelled on the Eiffel Tower.

that seeing would improve further during winter darkness when optical astronomy would be possible. An automated Australian telescope left at Concordia during the 2004 winter reported seeing conditions that seemed to rival those observed in space².

Star gazers

But not everything went according to plan for the Australians. Left unattended, in a bulbous green building known as 'the kiwi fruit', the instruments stopped working after a few months, and the group only obtained data up to 5 May, just one day after the beginning of the permanent winter night. Also the instruments were sensitive only to the seeing from 30 metres upwards. Nobody knew what was happening closer to the ground.

So this year, Agabi volunteered to test the Nice telescopes during winter. An Antarctic veteran used to working at temperatures of -40°C , Agabi discovered that winter was quite another challenge. Going outside, the crew risked instant frostbite. Faces and hands were worst off. As temperatures dropped below -70°C , face masks frosted over in seconds, but working without them was unbearable. "Everybody got burned," Agabi says. Fiddly tasks such as changing screws became interminable using thick mittens, yet the researchers could slip them off and work with glove liners for only a few seconds at a time. On a mild night, perhaps only -60°C , Agabi would venture outside to admire the sky and take photos. He felt confident that they were going to have a good winter.

But to Agabi's dismay, as winter darkness approached, the seeing seemed to grow steadily worse. "I told myself, that's normal, the atmosphere just needs to settle down," he says. However, by the end of May, after four weeks of complete darkness, he realized there was a serious problem.

How could the seeing be so poor when the Australian instruments had found it to be exceptional? Agabi launched weather balloons to measure the amount of turbulence from the ground upwards³. He discovered that around 90% of the turbulence at Concordia lay below

30 metres. "The Australians started just above the interesting part," says Aristidi.

Concordia's turbulence is caused by an extremely steep temperature difference between the snow and the air above it. The air contains scarcely any water vapour — the Earth's principle greenhouse gas — and therefore retains little of the Sun's warmth. So the switch from 24-hour daylight to total darkness has little effect on its temperature. The snow, however, a much better heat-retainer during summer, steadily loses heat during the winter darkness, cooling the air immediately above, and creating layers of air at different temperatures. Although the air at Concordia is extremely still, the slightest breeze ruffles the layers, making stars viewed through a telescope skip frustratingly from side to side.

Although Agabi was disappointed with his findings, they weren't yet disastrous; many telescopes are built on platforms 25 metres high. But to be sure that the layer stayed reliably low during the winter, Agabi needed to make continuous measurements. So in early July, with only five weeks of darkness left, Agabi persuaded the station manager to let



Balloons showed that a layer of turbulent air was disrupting observations of stars at Concordia.

him put a telescope on the roof of Concordia, at a height of 24 metres.

It was a difficult project. The roof slopes and the only access is a narrow spiral staircase with an overhead trapdoor. The Concordia team had to build a platform for the telescope in complete darkness and at temperatures of nearly -80°C .

Towering task

The results were worth it. At last, from its vantage point above the turbulence, the telescope recorded the exceptional seeing for which the astronomers had hoped. "The window was dirty but I have cleaned it," Agabi says. "And through it I've seen the true quality of the site."

Back in Nice, Agabi's colleagues are busy thinking of ways around Concordia's unexpected turbulence layer: a large windbreak or a new form of adaptive optics that might counteract the stars' skipping. At Concordia the wind comes mostly from the same direction, so a single barrier might prevent even the faintest breeze from stirring the air. If neither of these options work, the answer would be to build towers. "At 30 metres we'd have practically the same performance as in space," says Aristidi.

John Storey, from the University of New South Wales in Sydney, who is a member of the Australian team, agrees. "Thirty metres is a bit disappointing, but it's not a show-stopper," he says. "Most of the world's 4-metre telescopes are at 30 metres or so above the ground." His team has begun to model the aerodynamics of potential towers and snow-hills on which a future telescope could be sited.

Aristidi has already arrived in Antarctica to take over from Agabi in preparation for Concordia's second winter. With him are instruments to locate the turbulent layers more precisely, and to take infrared measurements. The cold temperatures and absence of water vapour mean that background infrared noise should be very low even in summer. A group from the University of Perugia in Italy hopes to have an infrared telescope installed by January 2007.

In the longer term, suggestions for big astronomical projects at Concordia are plentiful. One idea is to build an array of 50 small telescopes, which could look for Earth-like planets. The only rival technology at the moment involves space launches, although if every single telescope needs a high platform, the project could lose its appeal.

In that case, Agabi thinks that one very large dish might be the answer. Constructing a single tower would be easier, and they could transport large components overland. Either way, he says, it is important to think big. "It really is a super site," he says, "and it's going to be worth truly grand projects."

Gabrielle Walker is a science writer based in London.

1. EPICA members *Nature* **429**, 623–628 (2004)
2. Lawrence, J. S. et al. *Nature* **431**, 278–281 (2004).
3. Agabi, A. et al. *Pub. Astr. Soc. Pacific* (in the press).

K. AGABI

K. AGABI

Designs on life

Earlier this month, students from around the world locked horns in competition. Their challenge was to build functioning devices out of biological parts. **Erika Check** finds out how they got on.

Even if you're thinking big, you usually have to start small. Especially, as a group of Swiss students found, when big means counting to infinity. The team was drawing up a blueprint for the world's first counting machine made entirely of biological parts. Although they had their sights on loftier numbers, they opted to go no higher than two. If the plan worked, it would be a proof-of-principle for a much larger tallying device.

The group, from the Federal Institute of Technology (ETH) in Zurich, was one of 17 teams unveiling their projects at the first international Intercollegiate Genetically Engineered Machine (iGEM) competition, held at the Massachusetts Institute of Technology (MIT) in Cambridge on 5 and 6 November. The event attracted students from all over the world to design and build machines made entirely from biological components such as genes and proteins. They drew up grand designs for bacterial Etch-a-Sketches, photosensitive t-shirts, thermometers and sensors. And if none of the designs succeeded completely, that was more because of the limitations of the nascent science of synthetic biology than any lack of enthusiasm, creativity or hard work.

Synthetic biology aims to merge engineering approaches with biology. Researchers working at the most basic level are copying simple biological processes, such as the production of a protein from a gene. They break the process down into its component elements, such as a gene and the pieces of DNA and other molecules that control its activity. They then string these elements together to build a module they know will behave in a particular way — say, oscillate between producing and not producing a protein, or produce a protein that can switch another module on or off.

It is these kinds of components — oscillators and switches — that engineers order from suppliers and link together to build more complex electronic circuits and machines. Synthetic biologists are trying to develop a similar armoury of biological components, dubbed BioBricks, that can be inserted into any genetic circuit to carry out a particular function. Scientists at MIT have established a Registry of Standard Biological Parts, a catalogue of BioBricks that theoretically can be ordered and plugged into a cell, just as resistors and transistors can be ordered and plugged into electronic circuitry¹⁻³.

But it is hard to find scientists who are trained and interested in both biology and engineering to fuel the development of this new science. So, like true engineers, the founding synthetic biologists are trying to build their future colleagues



Bidding for glory: teams from the ETH in Zurich (top), Cambridge, UK, (bottom right) and Massachusetts at the first international Intercollegiate Genetically Engineered Machine competition.

from the ground up. To do so, they have commandeered a time-honoured engineering tradition: the student competition. The iGEM event began life as a project class for MIT students in 2003. Last year, it was thrown open to other US universities, and this year it went international. The organizers hope to attract 30 to 50 teams next year, including some from Asia.

Competitive culture

Much like the robot competitions that tap into students' desire to build something cool, the iGEM jamborees fire the participants' natural curiosity — hopefully encouraging biologists to learn something from engineers, and vice versa. "If you want to make something in this field, you can't just get some glue out and stick two cells together," says Randy Rettberg of MIT, who organized the competition. "You have to learn some biology to do it, and it's easy to do that during the competition because you know exactly why you're doing it."

This year, the teams presented an eclectic

selection of designs. Students from the University of Cambridge, UK, tried to make a circuit that could control the movement of *Escherichia coli* bacteria. They aimed to engineer the bacteria to contain a switch governing their sensitivity to the sugar maltose. With the switch off, the microbes would ignore the sugar. Tripping the switch would make the bacteria sensitive to the sugar and induce them to move towards it. In the end, the group — like almost every other entrant — had trouble completing assembly of its genetic parts in time.

Many of the other students also tackled problems related to bacterial communication and motion. The team from Pennsylvania State University designed a bacterial relay race, which it hoped would bring synthetic biology into the realm of sports — an innovation that won it an award for the Best New Sport at the end of the competition.

A team from the University of Oklahoma's Advanced Center for Genome Technology in Norman tried to exploit the sugar arabinose as

E. CHECK

an engine to drive bacterial motion. Teams from the University of Toronto and the University of California, San Francisco, built concepts for bacterial thermometers; and groups from Harvard, Toronto and Princeton designed bacterial illustrators and Etch-a-Sketches. Detection and sensing were also popular, with groups from Davidson College and MIT focusing in this area. And a lab at the California Institute of Technology tackled a problem raised at last year's event: designing biological memory.

Students from the University of Texas, Austin, demonstrated the world's first bacterial photography system. The team engineered a plate of *E. coli* so that they would respond to light and has since used the invention to take numerous photos, including shots of the group's adviser, Andrew Ellington (pictured)¹.

Piece by piece

As well as helping students to bridge the divide between disciplines, the competition gave them firsthand experience of life in the lab. All hit snags assembling their parts into coherent devices. It is still difficult to dissect the different genetic components of the circuits, stitch them together and get them to work in live cells. As Emanuel Nazareth from the University of Toronto reported, the students all learned one hard truth: "You can never allocate enough time for assembly."

This hints at a larger problem in synthetic biology. The field aims to build up a library of parts that can be interchanged in circuits with minimal effort. But that goal is not yet a reality as DNA sequencing and assembling technologies are still a bit too expensive and complex. "We're not organized at the community level around fabrication," says MIT's Drew Endy, one of the founders of the field and of the Registry of Standard Biological Parts. These hurdles need to be overcome before biological components will be as easy to deploy as their engineering counterparts.

Another learning experience was the reminder that, even stripped down to its basic components, biology can be complex and unpredictable. A team from the University of California, Berkeley, for instance, tried to design an entirely new way for cells to

Bacterial portraits: this picture of Andrew Ellington (left) is converted to art by a lawn of bacteria.



J. TABOR/UNIV. TEXAS, AUSTIN

communicate. This is a potentially important because it would enable cells to send and receive information, forming interlocking cell circuits instead of relying on simple gene circuits built in single cells.

The team hoped to exploit a natural method used by bacteria to exchange genetic information, known as conjugation. In this, two bacteria connect their respective cell walls together using a structure called a pilus. The group managed to trigger the conjugation response with synthetic circuits. But the bacteria turned out to be so eager to join up that they did so in huge bunches — and once they did, it was hard to separate them. "They don't really conjugate one at a time," said team spokeswoman Melissa Li. "They can go, but they can't stop."

The living end

Scientists in synthetic biology expect to stumble over these unanticipated quirks. But the pioneers admit there is still a basic question that the field hasn't yet answered: will synthetic biology actually work? As Rettberg points out, there are a lot of sceptics. "There's this big question, which is: can you build simple biological systems out of interchangeable parts and make them work in living cells?" he says. "We think you can, but there's a lot of people who think the other way and say biology is simply so complex you can't do it."

Given these concerns, it is no surprise that researchers have decided to keep things as

simple as possible, hence the ETH team's approach. As Robin Künzler said in introducing the project: "In the beginning we talked about counting to infinity, but we thought maybe we'd start by counting to two."

Despite this narrowed scope, the students faced a formidable challenge. They designed a series of devices: the first takes in a signal, which is then passed to an event processor. This processor — a genetic circuit — splits the signal into two components. A third circuit does the counting and generates a read-out.

To design and execute its plan, the Zurich team enlisted student biologists, engineers and computer scientists. Together they successfully designed and tested the event processor. But they hit a snag when building the counter: the company making the genetic sequences was unable to deliver them on time. Other teams encountered similar problems. Researchers in the field hope that the proliferation of suppliers — including firms starting up in Europe — will address that problem, if not in time for next year's competition, then for later down the road.

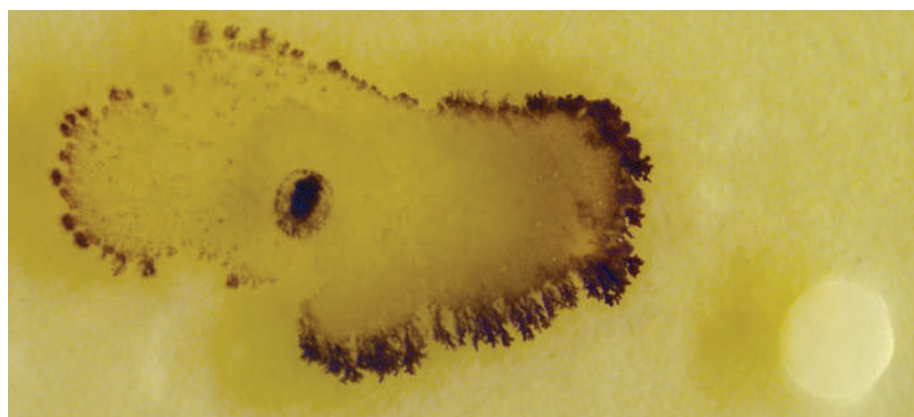
But iGEM has been as much of a learning experience for the old hands as for the students. "We don't know how to engineer biological systems," says Endy. "You can't teach something you don't know how to do, so the students are helping us to figure it out." After all, the field is young, and growing pains are inevitable.

Nevertheless, the plan to engineer future synthetic biologists seems to be paying off. Several institutes taking part in iGEM have confirmed that they intend to run classes in the subject as a result. "It's been a very interesting approach and what it has done is provide very rapid uptake and a very rapid spread of this whole idea," says Gos Micklem, a geneticist at the University of Cambridge. "The competition is essentially stimulating every level, from graduate and undergraduate to senior people."

Erika Check is Nature's Washington biomedical correspondent.

1. Levskaya, A. et al. *Nature* **438**, 441–442 (2005).
2. Sprinzak, D. & Elowitz, M. B. *Nature* **438**, 443–448 (2005).
3. Endy, D. *Nature* **438**, 449–453 (2005).

♦ <http://parts.mit.edu/wiki/index.php>



'Switched on' by an external trigger, *Escherichia coli* bacteria swim towards some maltose (bottom right).

BUSINESS

Property rights go East

China's approach to patents is undergoing a sharp transition, as **David Cyranoski** reports.

A stroll through the streets of Beijing or Shanghai, past the arrays of knock-off DVDs and Rolex watches on sale, gives many visitors a simple view of patent protection in China: they think there isn't any.

But behind the scenes, China is rapidly developing a strong, if unsung, commitment to intellectual property, according to a growing number of specialists in the field. And they say that companies in research-intensive industries that shun China for fear of patent theft could miss out on an upsurge in inventive, properly protected work there.

All of this was highlighted at a forum held in Beijing on 17–18 October to mark the opening of the Institute of Molecular Medicine of Peking University. Ian Harvey, chairman of London's Intellectual Property Institute, presented a roomful of scientists and pharmaceutical-company representatives with a surprisingly upbeat report on the situation.

"The legal infrastructure is among the best in the world," says Harvey, a former head of the British Technology Group, who has been visiting China for 30 years. The patent office there delivers patents of "good quality and reasonable cost", he adds. China's basic intellectual-property law has been updated several times since it was enacted in 1984, and the State Intellectual Property Office in Beijing has hired hundreds of examiners since the country joined the World Trade Organization in 2001.

Patents up

There is good reason for China to take intellectual property seriously — increasingly, it is producing its own. According to Harvey's data, Chinese universities applied for as many patents as US universities in 2002, and six times as many as Britain (see graph). Although the quality of these may be mixed, the boost in activity is undeniable. Paul Tam, vice-chancellor of the University of Hong Kong, says growing emphasis on intellectual property is transforming university research in China. "Patents are coming to be regarded as a sign of academic excellence," he says.

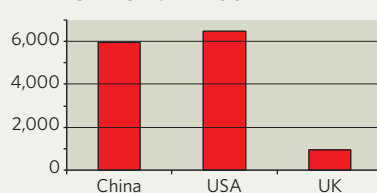
"There's nothing better for patent protection than the fact that there are now thousands of patents owned by Chinese citizens," says Robert Schaub, who heads the cardiovascular

and metabolic disease division of Wyeth, the Massachusetts-based drug company. Schaub, like many other drug-company representatives at the forum, is actively seeking opportunities for collaborative research in China.

The Achilles' heel of patenting in China is enforcement. Harvey admits that this remains a problem, with a major contrast between coastal cities such as Beijing and Shanghai, where he thinks courts are enforcing patent law, and parts of China's interior where they are not. But more courts are handling patent-infringement cases, and these are often speedy and cheap. Claims are typically dealt with in less than a year, Harvey says, compared with 18 months in Germany, two years in Britain, and five to six years in the United States.

Jiayang Yu, a partner at Beijing-based law firm Liu, Shen & Associates, says that enforcement is starting to work. In 2004 a Supreme Court ruling lowered the threshold for damages that allowed legal action for patent violations,

UNIVERSITY DOMESTIC PATENT APPLICATIONS IN 2002



making it easier to sue. More than 2,500 infringement cases were heard in Chinese courts that year, up 20% on 2003. "The increasing number of cases shows growing confidence in the system," Yu says.

But many companies remain unconvinced. Some do not apply for patents in China because this would require them to release information to potential counterfeiters, says Ralph Koppitz, a lawyer for Taylor Wessing in Shanghai who previously worked for the intellectual-property division of the European chamber of commerce in China.

As the patent-protection system comes into effect, Koppitz explains, this approach may prove expensive. "Many companies try to make a patent-infringement complaint," he says, "but then realize they never bothered to register in China," says Koppitz.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Production of knock-offs of drugs such as Pfizer's Viagra could soon wane, China-watchers say.

Sceptics sometimes cite the case of Viagra, Pfizer's drug for treating erectile dysfunction, as an example of lax patent protection in China. The patent was overturned by a Chinese court, and is now under appeal. Yet the fact that Pfizer sought to protect it as a second-use medicine — it was originally patented as a cardiovascular treatment — complicates the case, points out Steve Smith, a patent lawyer at NuPharm Intellectual Property in Holmes Chapel, UK.

Infrequent disputes

Smith thinks that, for firms in science-led industries such as pharmaceuticals, the spectre of large-scale intellectual-property theft in China is "much more a perception than a reality. When you ask people, they can't come up with examples."

Indeed, although the number of patent disputes in other industry sectors has risen, the drug industry has seen few of them. Yu explains that the cost of drug development and clinical trials is so high that most people would want to work with a drug that could win if challenged in court. "Would you invest US\$10 million in a company if it might infringe on someone else's patent? Probably not."

The protection of trademarks in China remains a serious problem — as does the counterfeiting of manufactured products, including drugs. Koppitz says that courts in provincial cities simply will not act against local manufacturers of knock-off goods. But newly researched ideas can be effectively protected,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

P. PARKS/AFP/GETTY IMAGES

and as high-tech companies realize this, they are contracting out more research there.

Beneficiaries of this trend include Shanghai-based WuXi PharmaTech, which was established only in 2001 but already makes drugs for 18 of the world's 20 largest pharmaceutical companies.

Theft of ideas is a major threat, concedes Ge Li, who heads the company. WuXi's plant is equipped with tight security measures. Material received from companies is made into digital copies and notarized so that the documents can be used to establish precedent in case of a patent dispute. "They only feel comfortable once they visit us," Li says of his customers.

But Li's best weapon is his willingness to use the legal system against errant employees. If someone in the company stole something related to intellectual property, "I'd go after him," he says. "Intellectual-property protection is the lifeline of this company."

Such dedication is winning over research-oriented companies. Last November, Roche announced the opening of a Chinese research unit, and Novartis followed suit this year.

This trickle could soon become a flood as firms respond to the growing scientific skill of Chinese laboratories and to an expanding Chinese market for pharmaceuticals, which grew in value by 28% to US\$20 billion in 2004, making it the seventh largest in the world. And the number of US patent applications from Chinese scientists and engineers grew from around 600 in 2002 to 1,800 last year. It is time to take China more seriously, declares Harvey. "It will become a major force — maybe the major force — in intellectual property." ■

IN BRIEF

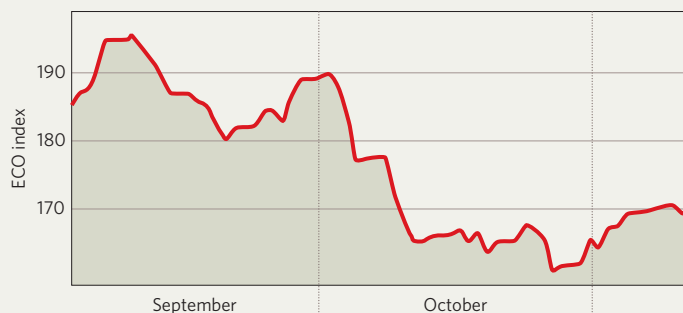
SUN RISING California computer maker Sun Microsystems has unveiled a chip, the UltraSPARC T1, which can process 32 threads of information at once. The Santa Clara-based company says that the chip consumes only 70 watts of electric power — less than half as much as some of its competitors. Computer-industry analysts say the chip, which will power large computer servers, breaks major new ground in terms of its parallel-processing power, but they point out that software compatibility issues might limit its success.

CHEMICAL SHIFT DuPont, one of the world's largest chemicals companies, has said it will shed 4,000 jobs in a major reorganization. No laboratory positions will be lost, the company says, but some research programmes, such as electronics and biologically based materials, will be given greater emphasis at the expense of others, including some polymer work. DuPont says the plan will save it \$2 billion over the next three years.

GENERIC GAIN The world's best-selling drug faces a looming generic threat after a company-sponsored study failed to prove that it is significantly more effective than a competitor in patients who have had heart attacks. High-dose Lipitor (atorvastatin), the \$12-billion cholesterol-lowering drug made by Pfizer, did not work significantly better than regular-dose Zocor (simvastatin), made by Merck, in preventing repeated heart attacks, cardiac arrest and deaths related to heart disease, *The Journal of the American Medical Association* reported last week (T. R. Pedersen *et al.* *J. Am. Med. Assoc.* 294, 2437–2445; 2005). The finding boosts the commercial prospects for generic versions of Zocor, which goes off-patent next June.

MARKET WATCH

CLEAN-ENERGY STOCKS



Clean-energy stocks have slipped back over the past two months after hitting record highs in the summer, when oil price increases to \$70 a barrel alerted investors to the potential value of stocks in alternative energy sources.

The Wilderhill Clean Energy Index — whose symbol on the American Stock Exchange is ECO — tracks the performance of companies whose businesses rely significantly on energy sources other than fossil fuels or nuclear power.

And as oil prices have fallen this autumn, the index has taken a hit. Nonetheless, says Robert Wilder, the former political scientist whose company runs the index, money continues to pour into funds set up to track it. "That's an indication of

strong, ongoing interest in the sector," he says.

Big movers in the past two months included California microturbine-maker Capstone Turbine, whose stock halved in value as the speculative interest that had boosted its price fourfold in the summer wore off.

However, stocks in solar-panel supplier Evergreen Solar of Massachusetts advanced by 50%, to almost \$12, on powerful global demand for its photovoltaic cells.

If oil prices fall back further next year, Wilder suggests, speculative investors may start to frown on alternative energy sources. But even with oil at, say, \$50 a barrel, he says the sector could enjoy sustained growth in demand for its products in the medium term.

Colin Macilwain

WILDERSHARES

Bush's policy stopped US gaining stem-cell lead

SIR — Your News story “Korea launches network to share cloning information” (*Nature* 437, 1077; 2005) reports the establishment of the World Stem Cell Hub, under the direction of Professor Woo Suk Hwang. Hwang's research team have developed a highly efficient recipe for producing human embryos through somatic-cell nuclear transfer (SCNT) and then extracting their stem cells (W. S. Hwang *et al. Science* 303, 1669–1674; 2004).

The announcement of the hub signals South Korea's intention to become the world's leading centre for stem-cell and therapeutic-cloning research. It also reflects how far the United States has fallen behind its competitors in this pivotal area and how much the lack of federal leadership has handicapped US efforts.

In 2000, Advanced Cell Technology (ACT) — of which one of us is vice-president for medical and scientific development — initiated a significant human therapeutic-cloning programme. ACT's ethics advisory board assisted researchers by providing ethical guidelines and supervision for a pioneering egg-donor research programme. In 2001, ACT scientists reported the creation of the first early (4–6-cell stage) cloned human embryos (J. B. Cibelli *et al. J. Regen. Med.* 2, 25–31; 2001). As early as 2002 and 2003, the team of researchers at ACT had very promising results — including what we believe were stem-cell-stage competent embryos — that seemed to be on a par with those of the South Korean team, subject to some changes in the experimental conditions.

Why did the South Koreans win this race despite our early lead?

In our view, President George W. Bush's restrictive policy on funding stem-cell research was a major factor. SCNT research is expensive — a full research programme costs hundreds of thousands of dollars each year. At that time, ACT was a privately financed company, and from the summer of 2001 on, it was operating in an extremely hostile funding environment, with no hope of federal support. There is no reason to believe that ACT was a special case. Indeed, the stem-cell area as a whole has continued to encounter difficulties in garnering sufficient financial support.

Bush also repeatedly spoke out in support of legislation in Congress that would ban all therapeutic-cloning research. Investors may be willing to accept market and research risks, but they are very reluctant to fund work that might be criminalized, and venture-capital funding dried up. By mid-2003, it had become a challenge for ACT to maintain staffing levels and meet payrolls.

In vitro fertilization clinics, too, were

unwilling to get involved. There was concern among clinic staff that they would receive adverse media publicity for participating in stem-cell research and that the physical security of staff and patients would be put at risk.

No one likes to lose a race. Apart from the egos involved here, however, the stakes for this research are important. Although the South Korean team deserve every credit for their accomplishments, the current absence of a strong US competitor in this research narrows the range of directions likely to be explored.

Robert Lanza*, **Ronald M. Green†**

*Advanced Cell Technology, 381 Plantation Street, Worcester, Massachusetts 01605, USA and Institute of Regenerative Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA

†Ethics Institute, Dartmouth College, Hanover, New Hampshire 03755-3500, USA

Evidence of group learning does not add up to culture

SIR — Andrew Whiten and colleagues' study of group learning, “Conformity to cultural norms of tool use in chimpanzees” (*Nature* 437, 737–740; 2005), is intended to show evolutionary continuity between humans and other animals, but the authors directly make that claim only for conformity.

Jacqueline Zupp, in Correspondence (*Nature* 437, 1089; 2005), says she finds it absurd to believe that human culture and society developed without precedent among animals and adds that we “now have evidence for animal cultures, as reported in the pages of this journal”.

The experiment by Whiten and colleagues is well conceived and executed, and is entirely convincing on its own terms. But continuity between learning to open a latch that is already there, and creating, say, *The Iliad* or the pythagorean theorem, is not obvious. Are we to value the proof above what is being proved?

The continuity of the human mind with the animal mind is the most important question in human evolution, so we want to get it right. But in our rush to triumph, we should not allow our conclusions to be driven tacitly by hints and implications and the use of emotional vocabulary in a way that would never be tolerated in, say, chemistry or mathematics. Like Zupp, I am an evolutionist, and I do not wish to see our science paint itself into a corner from which the only escape leads through a gauntlet of public embarrassment.

William L. Abler

Department of Geology, Field Museum, 1400 South Lake Shore Drive, Chicago, Illinois 60605, USA

Is the ID debate proof of an intelligent deceiver?

SIR — In the ongoing debate over whether intelligent design (ID) should be taught as a legitimate alternative to evolution in schools (“Expert witness: the scientists who testified against intelligent design” *Nature* 438, 11; 2005), I suggest that ID could be presented as an alternative so long as it is always accompanied by a third option: intelligent deception.

This hypothesis proposes that the ID movement is motivated by an ‘intelligent deceiver’. Individuals who understand how to debate alternative scientific hypotheses would never intentionally promote religious dogma as science. So an intelligent deceiver must be at work, guiding proponents of ID to sow confusion over valid scientific debate.

To exclude intelligent deception from debates over ID versus evolution could be considered hypocritical on both legal and moral grounds. And if proponents of ID reject the hypothesis of intelligent deception, their objections would be most interesting to hear, particularly the ones that dismiss the deceiver without imperilling the designer.

A. Richard Palmer

Systematics and Evolution Group, Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Librarians can help prevent accidental plagiarism

SIR — I agree wholeheartedly with your Editorial “Clamp down on copycats” (*Nature* 438, 2; 2005) stating that universities need to instruct students in the standards that are expected of them.

I have no doubt that some writers set out to deceive. But I wonder how much plagiarism takes place because people may not be aware of what they are doing? Librarians can make a real contribution to promoting good academic practice, by teaching about referencing and how to use reference management software. We can help academic and research colleagues in the rigorous instruction that you rightly advocate.

Keith Nockels

Clinical Sciences Library, University of Leicester, Clinical Sciences Building, PO Box 65, Leicester LE2 7LX, UK

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

COMMENTARY

Let us go forth and safely multiply

Synthetic biology, which involves the engineering of new biological components and organisms and the redesign of existing ones, will require community discipline and openness if it is to flourish safely, says **George Church**.

The developing field of 'synthetic biology' could be seen as yet another expression of scientific hubris. It has potential benefits, such as the development of low-cost drugs or the production of chemicals and energy by engineered bacteria. But it also carries risks: manufactured bioweapons¹ and dangerous organisms could be created, deliberately or by accident.

There is now a short window of opportunity during which the neologism can be attached to novel unions between the existing fields of genetic engineering, synthetic chemistry and metabolic engineering. Synthetic biology also needs to distinguish itself as a safe community effort that nurtures responsible practices and attitudes. For some, synthetic biology shares the potential, along with nanotechnology and artificial intelligence, to generate new entities that can reproduce and evolve at will^{2,3}. Whether we believe that these are immediate, distant or imaginary threats, the concerns are real. For their part, biologists are excited by the potential for manufacturing precise, reliable and scalable synthetic components.

Safety first

A code of ethics and standards should emerge for biological engineering as it has done for other engineering disciplines. The community recognizes this need, but discussions are fragmentary. The next international meeting on synthetic biology (in May 2006 at the University of California, Berkeley) should make significant progress in that direction. What practical guidelines ought to be considered?

First, proper use of physical-isolation measures, as is already prescribed by levels 2 to 4 of the biosafety laboratory standards. Level 2 requires biological safety cabinets, lab coats, gloves and face protection; level 4 specifies a separate building, full-body suits and more.

Second, biological isolation — engineering biological systems to reduce their viability outside the lab and factory — should become standard practice. Genetic strains can be designed to require essential nutrients that are unavailable in the wild. The genome can also be engineered so that the genes cannot function in other cells. For example, a sufficiently novel genetic code for protein synthesis, not based on the standard amino-acid code, would not be expressed or function properly if taken

up by natural cells or viruses. These safeguards should prevent genes for new toxins, allergens or pathogens from mixing and stably recombining with wild species.

The list of precautions is limited only by our creativity. Engineered cells could be programmed to self-destruct after a fixed time or on detecting an external signal. Similarly, engineered sequences can act as 'watermarks' for easier tracking. Genetic sequences that move around easily, such as transposons, can be removed from cells, thereby reducing undesirable genetic changes⁴. Other safety features we can imagine go beyond single-

"Learning from gene therapy, we should imagine worst-case scenarios and protect against them."

gene manipulations associated with conventional genetic engineering. For example, useful industrial microbes could be designed to have reduced mutation rates, except under specific lab conditions. Other microbes could be modified to detect the production of undesirable chemicals.

We should encourage young scientists to think constructively and build environments for sharing biological resources, with an emphasis on safety, as is happening in the International Genetically Engineered Machine (iGEM) programme. This event involves a growing number of universities (13 this year) sponsoring students to construct synthetic genetic systems (see page 417).

Above all, outreach is required. Genetically modified products, including crops and gene-therapy drugs, have been opposed for reasons that go beyond worries about scientific uncertainties. Citizens who will gladly take recombinant-DNA drugs (such as interferon, insulin and erythropoietin) are reluctant to eat foods containing even trace amounts of recombinant DNA. Can synthetic biology gain greater public trust? We should learn from past cases; in the case of foods generated by synthetic biology, for example, we need to recognize that stakeholders include not just the farmers, but their neighbours and grocery shoppers also.

Learning from gene therapy, we should imagine worst-case scenarios and protect against

them. For example, full physical isolation and confined lab experiments on human or agricultural pathogens should continue until we have data on a greater number of potential consequences — ecological and medical — of engineering such systems. Moving from closed-lab commercial projects to open-air systems will require appropriate experimental procedures and perhaps higher levels of biological isolation.

A watchful eye

In addition to a code of professional ethics for synthetic biologists, we need to watch for the rare cases when they transgress. This requires not just laws, but also monitoring compliance⁵. This could exploit government experience in the surveillance of illegal drugs and hazardous materials. In the commercial sector, monitoring systems could reveal suspect activities, such as labs requesting DNA that is related to potentially harmful biological agents. The purchase of precursor chemicals, nucleic acids, genes and designer cells could be screened against a pathogen database. However, automated monitoring will require cooperation by manufacturers⁶ and international coordination. Discussions about this have begun, including one funded by the Sloan Foundation⁷. But any actions that penalize the legitimate manufacturer or user are likely to backfire, and having laws without government-mandated surveillance will be ineffective.

Finally, the community needs to discuss the benefits of synthetic engineering to balance the necessary, but distracting, focus on risks. From now on, each small step towards engineering enzymatic pathways for cheaper pharmaceuticals, smart biomaterials and large-scale integrated genetic circuits should be celebrated.

George Church is in the Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

1. Alibek, K. *Biohazard* (Delta, New York, 1999).
2. Joy, B. www.wired.com/wired/archive/8.04/joy_pr.html (2000).
3. Kurzweil, R. *The Singularity Is Near: When Humans Transcend Biology* (Viking, New York, 2005).
4. Kolisnichenko, V. et al. *Genome Res.* **12**, 640–647 (2002).
5. Church, G. M. *Synthetic Cell Projects* <http://arep.med.harvard.edu/SBP/> (2004).
6. Aldhous, P. *New Scientist* magazine **188** (2525), 8 (2005).
7. *Study to explore risks, benefits of synthetic genomics* <http://web.mit.edu/newsoffice/2005/syntheticbio.html> (2005).

BOOKS & ARTS

All the fun of the fare

A selection of amusing morsels from the history of nutrition science.

Terrors of the Table: The Curious History of Nutrition

by Walter Gratzer

Oxford University Press: 2005. 288 pp.
£18.99, \$30**Marion Nestle**

Terrors of the Table is Walter Gratzer's third book on the theme of science as entertaining anecdote. Since his retirement in 2002 from King's College London as a biophysical chemist, he has taken on the less seemly side of the scientific enterprise: the rivalry, envy, delusion, self-deception and outright fraud sometimes committed by researchers ostensibly engaged in the search for truth. As he explained in a radio interview in 2002, his purpose is "to astonish, to instruct and, most especially, to entertain".

And what could possibly be more entertaining than the history of nutrition? This, Gratzer says, is "full of fascination and drama, for it encompasses every virtue, defect, and foible of human nature". To reveal the drama — and a taste of the science that underlies it — he tells some familiar stories. We meet James Lind, who used lemons to counter scurvy; William Beaumont and his fistula patient; Christiaan Eijkman, whose milled rice and chickens revealed so much about beriberi; and a great many others. The stories are replete with sex, religion, suicide, an occasional murder and a riveting cast of characters. Antoine-Laurent Lavoisier, beheaded at the guillotine in the French Revolution, left behind a "high-spirited" widow, who married the physicist Benjamin Thomson (later known as Count Rumford). Thomson promoted the nourishing properties of soup and invented the drip coffee pot. The chemist Justus Liebig, Gratzer tells us, was "a giant, often wrong, but seldom in doubt". Albert Szent-Györgyi, who won a Nobel prize for his work on vitamin C, had four wives, two of them 50 years younger than him.

Along with the gossipy details, Gratzer tosses in such diversions as a poem on food adulteration by G. K. Chesterton, and a menu from an 1868 London banquet that lists, among other delectables, "collared horse-head". For readers who might still be interested in the science, an appendix explains the Krebs cycle, glycolysis and inborn errors of metabolism.

The stories are amusing but do not seem to be organized within any particular historical

A taste of history: a Thanksgiving Day dinner in nineteenth-century New England.

or social context. If the book has an underlying message, it is that "the paladins of progress were always up against the forces of reaction". This message is especially evident in the final chapters, in which Gratzer confronts nutrition evangelists, miscreants, vitamin-mongers and the "higher quackery" of food adulterers, diet purveyors and cholesterol zealots. As a modern example, he points to the "hopelessly obfuscated question of salt intake", because its connection to hypertension (except in a "small minority" of individuals) comes down to epidemiology and theology. "There are believers and disbelievers," writes Gratzer, but "the sceptics have won the argument". Really? He cites a sceptical journalist but ignores research reviews from decidedly secular groups, such as the 2005 US Dietary Guidelines Advisory Committee, that continue to interpret "small minority" as significant and to conclude that populations would be healthier if they ate less salt.

Such casual documentation is the most troubling aspect of Gratzer's anecdotal approach. It seems as though he has read books on nutrition history and written his own Readers Digest-style synopses. Entire sections lack citations, as do many quotations. I loved a comment attributed to Sydney Brenner

— that he "long ago discovered the obesity gene: it is the gene that opens the mouth" — but the book gives no source.

Of Joseph Goldberger, the US Public Health Service officer who identified pellagra as a dietary deficiency, Gratzer writes: "Goldberger's first posting was to a quarantine station, responsible for inspection of ships arriving at the port of Pennsylvania." No footnote is evident but the 'further reading' list mentions Alan Kraut's *Goldberger's War* (Hill & Wang, 2003), which I happen to own. Kraut says "Goldberger's first post would be New York" in 1899. He continues: "On April 24, 1900, Goldberger was reassigned by Surgeon General Walter Wyman to the Reedy Island Quarantine Station at Port Penn, Delaware, the station responsible for inspecting ships seeking entry to the port of Philadelphia" (citing a letter from Wyman to Goldberger in 1899). Only the most exacting of scholars will care whether this posting was Goldberger's first or second, but it suggests that the details of Gratzer's book are best taken with a modicum of that theological salt.

Nutritional problems ranging from infant malnutrition to obesity affect vast segments of the world's populations, and are caused as much by political, economic and cultural factors as

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

THE ART ARCHIVE/CULVER PICTURES

by the lack or excess of nutrients and energy. The many determinants of nutritional status, and the difficulties inherent in figuring out what people eat from day to day, make studies in this field exceptionally challenging. The history of attempts to disentangle the effects of diet from other factors has much to teach us about the ways in which science and society interact. Gratzner's gentle mocking of human

frailties makes pleasant reading, but does little to explain the complexity of the issues or the intellectual demands placed on scientists who try to address them. If Gratzner's book does any good at all, it will inspire students to take up this challenge. ■

Marion Nestle is in the Department of Nutrition, Food Studies, and Public Health, New York University, New York, New York 10012-1172, USA.

Unearthing mammalian origins

Mammals From the Age of Dinosaurs: Origins, Evolution, and Structure

by Zofia Kielan-Jaworowska,
Richard L. Cifelli & Zhe-Xi Luo

Columbia University Press: 2005. 630 pp.
\$195

Timothy Rowe

Palaeontologists may recognize *Mammals From the Age of Dinosaurs* as an update of an earlier classic, *Mesozoic Mammals*, edited by Jason Lillegraven, Zofia Kielan-Jaworowska and William Clemens (University of California Press, 1979), a volume of tremendous and lasting impact.

But the new book, for which Kielan-Jaworowska was joined by Richard Cifelli and Zhe-Xi Luo, is twice the length of its predecessor. It is broader in content and scientific scope, and is in many ways a different book altogether. Its increased girth reflects recent areas of scientific advance, including exciting fossil discoveries from across the world, new technologies for analysing them, and changing philosophical perspectives on this defining segment of our own distant past. The book is extensively illustrated, comprehensive and detailed in its treatment of Mesozoic fossils and their primary literature. It presents a

broad historical synthesis of early mammalian evolution. And it is likely to become the first book to reach for when embarking on the trail of virtually any question or problem involving deep mammalian history.

The known fossil record of Mesozoic mammals has grown faster in the past 30 years than in the previous 300. Although overshadowed in the popular media by feathered dinosaurs, more than 200 new Mesozoic species have been named and designated as mammals, and they are profoundly reshaping our knowledge of mammalian structure and evolution. Most of the new names refer to isolated teeth and jaws, but some spectacular discoveries include complete skulls or articulated skeletons, and a few specimens even provide evidence of fur. The three authors of this volume have been directly involved in many of these discoveries, adding authority to this book.

The book overwhelmingly reflects the world view of George Gaylord Simpson, and shows signs of Willi Hennig's influence and the importance of phylogeny reconstruction in interpreting history. But as Isaac Newton proved in deriving the calculus, two variables cannot be optimized in a single equation. The book's architecture thus reflects a struggle that has dominated palaeontology for three decades.

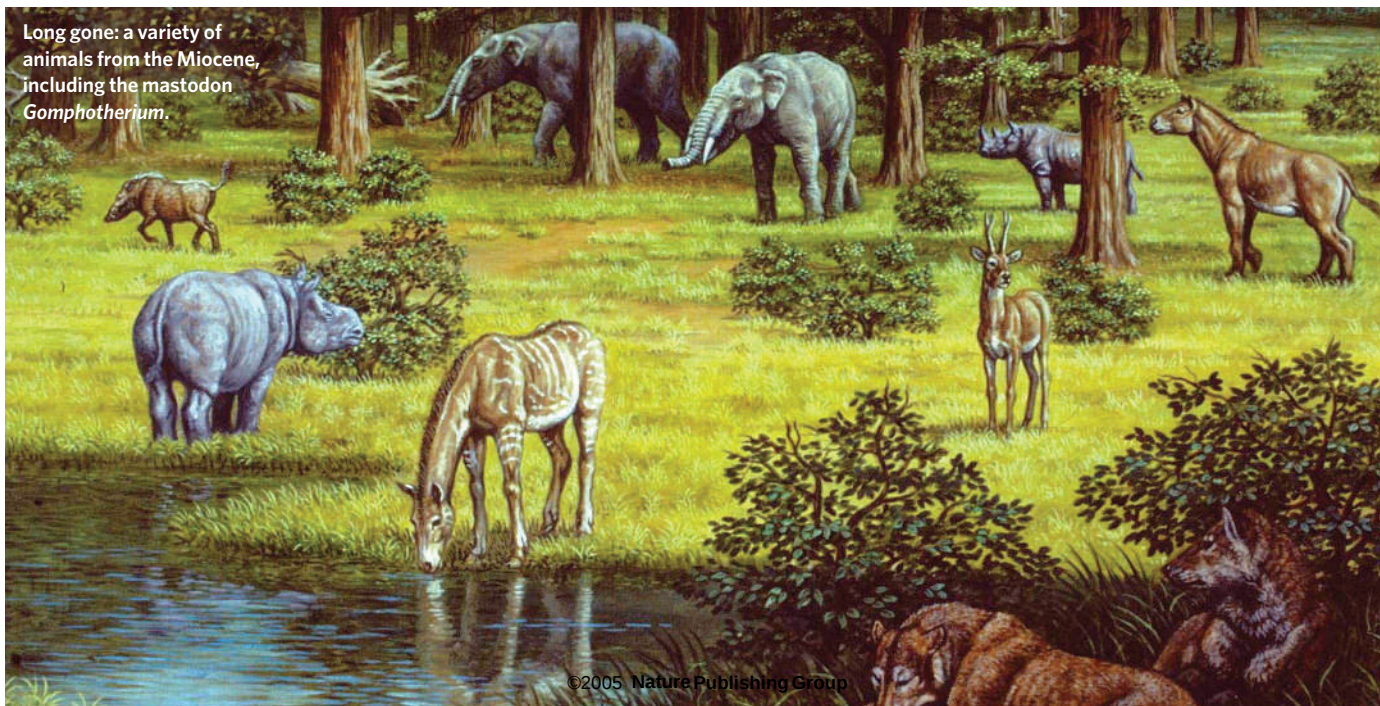
By giving primacy to a simpsonian view, *Mammals From the Age of Dinosaurs* presents an encyclopaedic work whose thoroughness ensures it will have long and useful life. But its phylogenetic framework is important to its overarching historical synthesis, and that framework was challenged in several papers published by Michael Woodburne and colleagues as this book was nearing completion. Taxon selection, character conceptualization and polarity, and other technical issues will be debated by those in the field as long as this book is read. If the critics are correct, then major structural features of the historical argument may crumble as well. The most important is a fundamental division of mammals into Australosphenidans and Boreosphenidans, and this casts doubt on the authors' interpretations of character history, such as their views on the dual origins of the tribosphenic molar and mammalian middle ear. But even if criticisms of this book's phylogenetic architecture are upheld, it simply highlights that much is yet to be learned about mammalian phylogeny and history.

This beautiful book surpasses all its predecessors in its scope and detail, and in using the most rigorous methods yet directed at such a broad panorama of early mammalian history. Its informative narrative attests that mammalian palaeontology is in the midst of a renaissance of discoveries unexpected 30 years ago. The rich resources gathered between its covers are timely and exciting, and should foster a further boom in mammalian palaeontology.

I am glad to own two copies of this book. It will be the basic mediator of debate for many years to come, and I expect to wear out both copies long before a comparable work emerges to overshadow it. ■

Timothy Rowe is at the Jackson School of Geosciences, and Texas Memorial Museum, University of Texas at Austin, Austin, Texas 78712, USA.

Long gone: a variety of animals from the Miocene, including the mastodon *Gomphotherium*.



A poisonous present

Kampfstoff-Forschung im Nationalsozialismus: Zur Kooperation von Kaiser-Wilhelm-Instituten, Militär und Industrie. [Weapons Research in National Socialism]

by Florian Schmaltz

Wallstein: 2005. 676 pp. €39

Benno Müller-Hill

Germany started developing chemical weapons nearly a century ago, during the First World War. Fritz Haber, then director of the Kaiser Wilhelm Institute of Physical Chemistry in Berlin, collaborated with the chemical industry and the army to set up Germany's powerful industrial-military complex (see *Nature* 438, 158–159; 2005). After the war, research on chemical weapons in the Weimar Republic was forbidden by the Allies, but it was still done secretly on a small scale. In *Kampfstoff-Forschung im Nationalsozialismus*, Florian Schmaltz tells how this research re-emerged under the Nazis between 1933 and 1945.

When Hitler came to power in January 1933, Germany no longer had a centre for research on chemical warfare. Haber was still director of the Kaiser Wilhelm Institute but he was told that his group leaders must be fired because they were Jewish. Haber, who was also Jewish, decided to go too, leaving the institute in ruins. Three young chemists from Göttingen, all members of the Nazi party, were sent to Berlin to take over.

Gerhart Jander, an inorganic chemist, had friends in the SA (the stormtroopers) and was forced to leave after the Röhm putsch, when Hitler purged the SA. Rudolf Mentzel, an inexperienced chemist, had no interest in doing research himself and became a science administrator in the Reichsforschungsrat, a German funding agency. In 1935, Peter Adolf Thiessen, a physical chemist, became the institute's director. The research focused on an explosive known as N-Stoff (chlorine trifluoride), which Thiessen hoped would prove more destructive than nitroglycerol. Despite being produced in large amounts, it was never successfully used. After the war, Thiessen made a career as a research administrator in East Germany, although Schmaltz does not discuss that here. He does, however, discuss five other Kaiser Wilhelm institutes that worked on gas warfare. Most were occupied with perfecting gas masks, clothing and shoes as protection against weapons such as mustard gas.

The most disquieting part of the book is the story of Richard Kuhn, who won a Nobel prize in 1938 for his work on carotenoids and vitamins. One-third of his lab at the Kaiser Wilhelm Institute for Medical Research in Heidelberg was occupied with research on poisons that could be used as weapons. When his colleague Otto Meyerhof left the institute

in 1938, Kuhn took over some of his lab space. This gave him access to the nerve poisons tabun and sarin, synthesized in secret in 1936 and 1939 by chemists of the IG Farben companies. The Allies only learned of their existence shortly before the end of the war. Kuhn's colleagues showed that the poisons worked by inhibiting acetylcholine esterase, and his laboratory synthesized an even more effective acetylcholine esterase inhibitor, soman, in 1944. These poisons were synthesized in large quantities for use in grenades, but they were never used. One reason is that it was impossible to protect German soldiers and civilians from the poisons. According to Schmaltz, "these poisons were a present for the future".

Kuhn was also a great science administrator. As a member of the Reichsforschungsrat, he chose to finance the mustard-gas experiments of his colleague Otto Bickenbach, who proposed to test vitamin B₆ as a possible protective agent against the gas. For his experiment, he mainly used gypsies from the concentration camp of Natzweiler-Struthof, four of whom

died. Bickenbach was later condemned by a French court to 20 years in jail. During the lawsuit, Kuhn wrote to Bickenbach's lawyer, saying that the experiments were "scientifically perfect" and "a blessing for future generations". Bickenbach was released in 1955 after a German medical court found nothing wrong with his experiments and allowed him to practice medicine again.

Schmaltz, a historian, presents all this and much more in great detail. The flow of money between industry, the army and science, in particular, is well documented, but the chemistry is flawed in places. The structures of sarin and soman are given, but that of tabun is missing, and the structure of pinacolone alcohol, a precursor of soman, lacks an OH group. In addition, not everything presented in the book is new. And the author also acknowledges that some documents may have been destroyed or lost in Russia, so the story is not complete.

The book should appeal to all those interested in chemical warfare or in the history of the Kaiser Wilhelm institutes — but it may be too detailed for the general reader. ■

Benno Müller-Hill is professor emeritus at the Genetics Institute, University of Cologne, Cologne 50674, Germany.

The latest on latex

Tears of the Tree: The Story of Rubber — A Modern Marvel

by John Loadman

Oxford University Press: 2005. 336 pp.

£19.99, \$37.50

Robert W. Cahn

In 1770, an artists' supply shop in London offered for sale half-inch cubes of a mysterious material that was called 'rubber' because it could be used to rub out pencil marks. The price was 3 shillings per cube, a large sum at the time — but then, nothing else was as effective in correcting artists' mistakes.

This is one of many pieces of information offered in John Loadman's miscellany of a book, *Tears of the Tree*. The book's purpose, he announces at the outset, is "to examine the story of natural rubber in its social context". In fact, about half the book is devoted to the social context, and this is its main merit.

The story of rubber begins with its discovery and use by Meso-American civilizations in what are now Guatemala and Mexico. In particular, rubber balls were used for formal competitive games; the losers were usually killed, their heads cut off and coated with latex from the rubber tree to make the next (properly weighted) balls. But the most horrendous stories in the book are about the 'rubber barons' in the western extremes of Amazonia, and the exploitation by Belgian King Leopold II of the natives of the Congo in the late nineteenth

century. The social context of primary rubber production was uniformly depressing until 1877 when, after numerous abortive attempts, seeds of the wild rubber tree were at last successfully (if unofficially) transferred from Brazil to the Royal Botanic Gardens at Kew, near London. They were then germinated and transferred to the Far East, Malaya in particular, establishing a major primary industry.

The exploitation of rubber provided extensive benefits for developing industrial societies. The story — especially the slow, faltering steps to find the best way to harden raw rubber by vulcanization — is told in interesting detail. Patents play a major role in this tale, and the biographies of the principal players are excellently presented. The cycle is completed with a critical survey of attempts to recycle waste rubber.

The book then changes gear and focuses on the production technology of vulcanized variants of rubber, both natural and synthetic. The treatment becomes extremely technical in terms of the organic chemistry of different kinds of polymeric molecules, and even more so when the minutiae of the chemical deterioration of rubber are set out. At this point, in my view, the book begins to stutter.

The science and technology of polymers in general, and of elastomers in particular, has three essential constituents: physics, chemistry and production technology. With a background in chemistry and long experience in a

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Tapping into nature: the production of tyres and medical gloves starts here, at the rubber tree.

research institute focused on natural rubber, Loadman knows about chemistry and production technology, and his presentation of the chemical aspects is technically detailed. But explaining the chemistry is not enough.

The problem comes with the physics. Load-

man claims that the physical theory of 'rubber-like elasticity' is too involved to present in any detail, so the source of rubber's elastic restoring force and of the very large elastic strains that can be attained is left unexplained. The elastic behaviour of rubber is fundamentally different from that of other, non-polymeric, solids: it is entropy made tangible, and no explanation is adequate unless it takes into account the probabilities of different microstructural arrangements of the polymer chains.

One difficulty in the book is the author's use of the word 'elastic'. He uses it to denote the capacity to generate very large reversible strains — that hasn't been the sole meaning since Hooke's work on elasticity in the seventeenth century. Furthermore, rubber, especially before it has been vulcanized, is not elastic but viscoelastic: the strain generated by an applied force varies with time under stress. This characteristic governs many features of rubber, including the adhesion of a tyre to the road surface and the comfort of someone sleeping on an elastomeric mattress.

The book is full of interesting information. But it is uneven, and the gear changes are so abrupt that the narrative tyre tends to skid. ■ Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

INSTALLATION

Uranium days

The Children of Uranium

At the Museo d'Arte Contemporanea di Villa Croce, Genoa, Italy, until 18 December. In English. www.museovillacroce.it

Sylvie Coyaud

Peter Greenaway is best known as a filmmaker. But he actually trained as a painter, and his particular sense of aesthetics is exhilaratingly apparent in his new multimedia operacum-installation, *The Children of Uranium*, which opened during the Genoa Science Festival earlier this month.

Created together with theatre director Saskia Boddeke, the work explores Greenaway's longstanding concern with the history of the element uranium. It is played out in a nineteenth-century palazzo, now used as an art museum. The palazzo's first floor and the grandiose staircase leading to it were cleared and redecorated for the installation. Each of the eight rooms is home to a different protagonist — from Marie Curie to George W. Bush — all of whom have played some role in society's

evolving relationship with uranium.

The audience is free to wander between the different sets. Their senses are bombarded by a mix of science, history, politics and philosophy during 90 minutes of song, dance and recitation. The characters of the nine performers change according to the environment.

In the lobby, Isaac Newton shaves, builds something — perhaps a prism? — using wooden tools, meets a naked Eve with whom

he shares, in dance, an apple, which is suspended by an invisible thread from the ceiling. The music rises and a tenor climbs the staircase singing, in a slightly threatening chant, the names of the first 92 elements of Mendeleev's periodic table. Despite his eighteenth-century court costume, this is the angel Moroni, who led the founder of the Mormon Church, Joseph Smith, to Utah and to the uranium mines of Moab. Upstairs he becomes Smith and ushers the audience into a room made up as the Oval Office, where President Bush plays with pencils, mulls over how to destroy the axis of evil, gets bored and practises his golf shots.

Meanwhile, in another room, Marie Curie kneels on pitchblende pebbles, surrounded by cabinets of laboratory glassware, and talks to her dead husband. "What have I obtained, what have I lost?" she cries over and over again. Butterflies seem to fly out from the pattern of the wallpaper to follow her in a dim beam of light, in a scene reminiscent of Gabriel García Márquez's *A Hundred Years of Solitude*, vanishing as she enters Robert Oppenheimer's grey cave. Here Oppenheimer watches, in anguished silence, the first nuclear explosion, projected in a continuous reel on to a fine-meshed screen, as if on smoke. He clutches to his chest a framed picture of a deformed baby.

Next door, Albert Einstein writes, smokes a pipe and chats with other performers in a study cluttered with books and papers, blackboards and equations. His letter of August 1939 to Franklin D. Roosevelt, in which he urges the president to build an atomic bomb before the Nazis do, is projected on the floor.

In a prison cell, Nikita Khrushchev reenacts his speech at the United Nations during the Cuban missile crisis, while the speech is shown on a television set. Mikhail Gorbachev mourns his wife, Raisa, next to her coffin. Her dresses hang listless from the ceiling, from which water starts to leak.

"Political decisions should not be left to scientists," sing the tenor and the soprano. Were they ever? But, the opera seems to tell us, perhaps they should also not be left to the bored

US president, nor to Khrushchev, drunk on vodka and the power of his nuclear weapons, nor yet to Gorbachev, a loser, broken by the death of his wife and the fall of the Soviet Union. "We are all the children of uranium," the singers continue. How we use it is for us, the people, to decide.

The live performance comes to an end, but recordings have been intelligently incorporated into the installation. Its large scale and small details challenge and perplex. Greenaway's relationship with physics remains ambiguous, but his creation is visually stunning. ■ Sylvie Coyaud is a science journalist based in Milan, Italy.



Albert Einstein makes the case for the bomb in *The Children of Uranium*.



D. ROBERT & L. FRANZ/CORBIS

Pushing for power

Tales of brilliant scientists and their heroic discoveries can overshadow the true nature of scientific communities, which are often dominated by battles for power and success.

Ad Lagendijk

Every branch of science cherishes its heroes. And admiration for them extends well beyond each discipline; laymen share the fascination when they learn about the lives of these champions of science and their brilliant discoveries.

Knowledge about these icons typically comes in the form of short, idealized biographical sketches which are supposed to reveal the context of scientific discoveries. They can be found in textbooks, newspaper articles, science magazines and scientific presentations. But the historical overviews of physicists and the physics community, seen through rose-tinted spectacles, strongly contrast with my daily experience as a professional physicist.

When I participate in a scientific conference I see a gathering of aggressive men (and yes, I mean men) fighting for their scientific claims to, at best, minuscule advancements. Territorial behaviour emerges all over the place, and yes, I too am guilty of it.

Successful scientists incessantly travel around the world performing their routines like circus clowns — forcefully backing up assertions over what are their contributions to the latest scientific priorities. Recently, there has been a call for physicists to focus more on biology. But surely there is no further need for this; physics 'red in tooth and claw' is already dominated by biology, of the kind studied by Charles Darwin.

A modest Japanese presenter does not stand a chance against a loud, American critic speaking in his, and modern science's, mother tongue. An offensive question asked at a conference by a streetwise, senior physicist of an overenthusiastic, junior Spanish scientist can be counted on to have the

desired effect: a high-tempered, ultra-fast, absolutely unintelligible reply. 'Target neutralized' as they say in the military.

It is not just at conferences that predatory scientists participate in the power game. Other forums include harsh reports written by anonymous referees reviewing papers for high-impact journals; damning assessments of a lecturer's teaching skills; or dismissive reviews of applications to granting organizations.

Powerful programme-committee members regularly appoint themselves as invited speakers. Even in what has been traditionally small-scale science, we physicists cultivate the mushroom approach: there is a rampant rise in the number of scientists running larger and larger groups, collecting authorships, citations and invitations to talk at international conferences.

A new quantifier of scientific power — number of patents — is already creating a furore. Patents will soon be added as trophies to the collections of these 'operators'.

In the battle for tax-payers' money, criticizing other branches of natural science, or indeed neighbouring disciplines in physics, is already a popular activity. "My discipline is more fundamental than yours," is a frequently heard claim. Note that it is not size that matters, but fundamentality. This silly, 'fundamentalist' game — reminiscent of high-priests fighting over the exegesis of their holy scriptures — is a favourite pastime for high-energy theoretical physicists. But all physicists enjoy this sport when commenting on chemistry or biology.

Young, self-assured, male PhD students quickly learn the rules of the game. When confronted with a new research assignment, their response is not fascination or curiosity; rather their first question is

whether they will be first author on resulting publications.

I am sure that my observations extend to other branches of science as well. Surely their practitioners similarly try to uphold idealized fairy tales about their discipline and its heroes for the public.

The primitive value system — tallies of publications, citations and patents — now used in science is the cause of this obsession with power rather than with curiosity and scientific progress. But does this system give us, in the long term, the best value for money? I doubt it.

A cynic will be ready with a response to my criticism: "Existing in the physics community is part of life and life is tough. Expecting higher moral behaviour of scientists is naive. Just read Thomas Kuhn."

But I have a different opinion. Science has always been a man's world. The values and norms that control our disciplines were established by men. In physics there is an alarming lack of female participants; it would be tempting to claim that because of physicists' typically masculine power games the physics community is not an attractive option for female scientists.

I won't make this claim, as the widely recognized, severe under-representation of women in physics has been analysed thoroughly, and from many angles, and nobody has found an easy solution. But I will say that to bolster true scientific progress, we should change our norms and values in physics. We should become emancipated. ■

Ad Lagendijk is group leader at the FOM-Institute for Atomic and Molecular Physics, Amsterdam, university professor at the University of Amsterdam, and professor of physics at the University of Twente, all in the Netherlands.

QUANTUM COMPUTING

A bit chilly

Leonard J. Schulman

A quantum computer needs a constant supply of 'qubits' in a known state. A nuclear magnetic resonance experiment that cools qubits by pumping entropy into a heat bath is a step closer to that goal.

On page 470 of this issue, Baugh *et al.*¹ demonstrate progress on one item in a long list of requirements for a functional quantum computer^{2,3} — ensuring a continuous supply of the basic carriers of quantum information (quantum bits, or qubits) in known states. The experiment was small-scale, involving just three qubits and slight cooling to prepare them, but it potentially shows the way to preparing larger numbers of qubits in well-defined states.

The possibility that quantum-mechanical effects might be used to speed up computation was little expected and long overlooked. The origin of the possible advantage lies in a strange feature of quantum theory: whereas the state of a classical system with N members can be specified by a number of parameters that are linear with regard to N , a quantum system needs a number that is exponential in N .

In 1982, Richard Feynman noted that this discrepancy has implications for computing⁴: on the one hand, a seemingly unavoidable exponential deceleration in classical simulations of quantum systems; and on the other, potentially exponential accelerations on quantum computational devices. A quantum algorithm that was demonstrably faster than its classical counterparts⁵ was developed in 1993; one year later, a quantum algorithm was found that could crack the RSA cryptosystem, the leading method for encrypting and authenticating communications⁶.

More than a decade on, RSA is still used to protect electronic commerce. How can this be? It turns out, simply, that quantum computers are very hard to build. What is required is a highly controllable, many-qubit device that can be maintained in complex quantum states over long periods of time. Such a device is too extraordinarily fragile to exist under normal circumstances — which is why no one has ever seen Schrödinger's infamous cat, dead or alive. A successful quantum computer needs no cat; but it must, like the cat, embody the predictions of quantum theory at a scale of size and complexity at which that theory has not yet been verified.

Baugh and colleagues' advance¹ comes in a rather more modest system, and exploits one of the better-explored methods for supplying

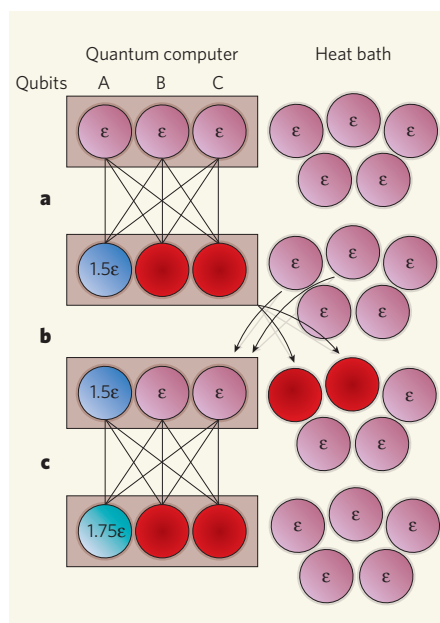


Figure 1 | Draining the heat bath. The heat-bath algorithmic cooling used by Baugh and colleagues¹ to prepare a quantum bit in a cooled state. **a**, After initial preparation, the three qubits of the quantum computer and many more in an ambient heat bath have a uniform bias ϵ to their lower-energy ground state. The overall bias of the three-qubit system can be shown¹⁰ to be the summed bias of all three qubits divided by two (1.5ϵ). One of the qubits (A) can be cooled by making it inherit the higher bias of the three-qubit system. The excess entropy is taken up by the other two qubits (B, C). **b**, These qubits B and C can be swapped for qubits with the original bias from the heat bath. **c**, In a future step, the procedure from **a** could then be repeated, with qubit A inheriting the new overall system bias (1.75ϵ). A limit of 2ϵ is approached after many steps.

and controlling qubits — nuclear magnetic resonance, or NMR. In this approach, each qubit corresponds to the two possible states of the spin of a nucleus — up and down. The first difficulty in using an NMR quantum computer is putting each qubit in a known state (initialization). The usual approach to doing this is to apply a strong magnetic field, the effect of which is to make the spin-up state a higher-energy, excited state, and spin-down a

lower-energy, ground state. This causes a preponderance of ground over excited states. The process extracts entropy from the system, and so can be thought of as a form of cooling.

However, the effect of this direct cooling is small: the 'bias' of a qubit (the excess probability of its being in the ground state over the excited state) is increased from 0 to a mere 1×10^{-5} . Although such an increase is good enough for the scientific and medical applications of NMR imaging, a general-purpose quantum-computing algorithm requires much better initialization⁷ — it needs so-called 'ancilla' qubits whose state is almost definite (having a bias close to 1). For useful computations, hundreds of these ancillas (at least) are required: direct cooling is, in short, an inadequate entropy pump.

The above difficulty is particular to NMR-based qubit applications, but a second difficulty is universal to all methods: a supply of ancillas is needed not only at the start, but also throughout any long computation. During a computation, the environment will occasionally interact with the device and cause coding errors. Fault-tolerance mechanisms that can prevent these errors from spoiling the computation do exist, but require additional ancillas.

The potential resolution to the difficulties of initialization and ancilla supply is algorithmic as much as it is physical. Qubits biased by just 10^{-5} towards their ground states may be thought of as highly disordered, or 'hot' (yet not infinitely hot — that would be zero bias, with total disorder between ground and excited states). A refrigeration process can be used to transfer entropy among a collection of qubits, cooling some while warming others. Eventually, the coldest qubits (those with the highest probability of being in the ground state, with a bias near 1) will be ready to be used in a computation. Such a refrigerator works at the level of individual qubit states and so can be regarded, and designed, as a computational algorithm.

A basic algorithmic cooling step was described by John von Neumann in 1956, albeit for a different purpose — reliable classical computation⁸. He calculated that if each of three bits has bias ϵ towards the 'right' answer to some question, then the bias of the system

as a whole (the 'majority vote') is given by approximately 1.5ϵ — so the bias of the whole system is greater than that of any individual qubit. The majority function can be computed 'reversibly' by a permutation of the eight possible sequences of the three bits (000, 001, ..., 111) such that one of the bits inherits the higher bias of the whole system and so is colder than it was before. The other two bits then pick up the excess entropy (Fig. 1a).

Closed-system experiments implementing this bias amplification have been performed previously using NMR^{9,10}, but the need for continuous ancilla resupply precludes a closed-system solution on a large scale. Instead, entropy must be pumped out of the computation qubits into an ambient heat bath¹¹, an approach known as heat-bath algorithmic cooling (Fig. 1b). What Baugh *et al.*¹ have achieved is an experimental demonstration of such a procedure. They take three hot qubits from a heat bath, put them in three computation nuclei, and then amplify the bias of one of these by a factor of 1.48 — remarkably close to the theoretically predicted performance of 1.5.

This is a significant first step along what will surely be a long experimental journey, leading, it is to be hoped, to heat-bath algorithmic cooling that can reach far lower temperatures on many more qubits. It may also be possible, even in the short term, to test a crucial aspect of open-system cooling: the use of continuous entropy pumping to achieve temperatures lower than those possible in a closed-system device. Specifically, in a closed three-qubit system, bias cannot be amplified by more than a factor of 1.5, whereas in an open system, a limit of 2 can be approached¹² by repeatedly recomputing the majority after exchanging the two 'used' (high-entropy) qubits for fresh ones from the heat bath (Fig. 1c). The modest gap between these two limits is a precursor of a much larger separation in many-qubit devices: a three-qubit open-system experiment exceeding an amplification of 1.5 will therefore be a notable milestone. ■

Leonard J. Schulman is at the California Institute of Technology, 1200 East California Boulevard, MC 256-80, Pasadena, California 91125, USA. e-mail: schulman@caltech.edu

CELL BIOLOGY

Silenced RNA on the move

Ralf Dahm and Michael Kiebler

Proteins are often produced at their site of action, but the RNAs from which they are made must be kept inactive until they reach the right spot. It seems this 'silencing' of RNA is linked to its transport around the cell.

Almost all cells possess 'compartments' that enable them to separate different biological tasks. Neurons, for example, have dendrites that receive input from other cells, axon hillocks that integrate the information, and axons that transmit signals to other neurons. A key mechanism that cells use to create this functional subdivision is the localization of specific messenger RNAs (mRNAs) to distinct cellular domains¹, which allows the cells to fine-tune gene expression in both space and time. But for mRNA localization to divide up a cell effectively, synthesis of the encoded protein must be repressed until the mRNA arrives at its site of action², otherwise the protein will be made and begin to act all along the journey³. This strongly implies that there must be a tight coupling between RNA transport and repression of protein translation. In this issue, Hüttelmaier *et al.* (page 512)⁴ provide the first evidence of a direct link between the two processes.

The authors studied the localization of β -actin mRNA in migrating cells. During their journey, cells extend protrusions in front of them to explore their path. These projections are generated by actin proteins polymerizing into long filaments that push the cell's membrane outwards. The β -actin mRNA is transported to the leading edge, so that the huge amounts of actin required for filament growth can be produced quickly and locally to establish cellular asymmetry (or polarity). Fibroblast cells, which proliferate around wounds and move inwards to fill in the lesion crater, are a classic example of migration (Fig. 1). This migration behaviour can be conveniently exploited in cell culture to dissect the

molecular mechanisms underlying the polarization and directed migration of cells.

Previous work from the same laboratory⁵ identified a short nucleotide sequence (the 'zipcode element') that is necessary and sufficient to move β -actin mRNA to the protrusions of fibroblasts and other migrating cells including neuroblastoma cells. The protein ZBP1 binds to the zipcode sequence, and is essential for this localization in fibroblasts and for the formation of dendritic filopodia, the precursors of synapses (the contact points between neurons)⁵. Now, Hüttelmaier *et al.*⁴ demonstrate that ZBP1 also represses the translation of β -actin mRNA, both *in vitro* and in intact neuroblastoma cells. Moreover, neuroblastoma cells lacking functional ZBP1 do not repress the translation of β -actin. However, when ZBP1 is reintroduced into these cells, translation is again repressed. Together, these experiments show that RNA transport and translational regulation are more intimately linked than previously anticipated.

But what occurs to switch the mRNA from the translational 'off' state that prevails during transport to the 'on' state once the RNA-protein complex reaches its destination? A close inspection of the ZBP1 sequence revealed a potential SH3-binding motif. Such motifs can serve as docking sites for the non-receptor tyrosine kinase Src — an enzyme that adds phosphate groups to other proteins. Hüttelmaier *et al.* show that Src does indeed phosphorylate ZBP1 (on tyrosine 396) *in vivo*, and that the phosphorylation makes ZBP1 much less able to bind to β -actin mRNA *in vitro*. Similar mechanisms regulate the

1. Baugh, J., Moussa, O., Ryan, C. A., Nayak, A. & Laflamme, R. *Nature* **438**, 470–473 (2005).
2. Gottesman, D. presented at Joint IPAM/MSRI Workshop on Quantum Computing, Los Angeles (2002).
3. DiVincenzo, D. P. *Fortschr. Phys.* **48**, 771–783 (2000).
4. Feynman, R. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
5. Bernstein, E. & Vazirani, U. *SIAM J. Comput.* **26**, 1411–1473 (1997).
6. Shor, P. W. *SIAM J. Comput.* **26**, 1484–1509 (1997).
7. Ambainis, A., Schulman, L. J. & Vazirani, U. in *Proc. 32nd ACM Symp. Theory of Computing* 697–704 (ACM Press, New York, 2000).
8. von Neumann, J. in *Automata Studies* (eds Shannon, C. E. & McCarthy, J.) 43–98 (Princeton Univ. Press, 1956).
9. Sørensen, O. W. *Prog. NMR Spectrosc.* **21**, 504–569 (1989).
10. Chang, D. E., Vandersypen, L. M. K. & Steffen, M. *Chem. Phys. Lett.* **338**, 337–344 (2001).
11. Boykin, P. O., Mor, T., Roychowdhury, V., Vatan, F. & Vrijen, R. *Proc. Natl Acad. Sci. USA* **99**, 3388–3393 (2002).
12. Schulman, L. J., Mor, T. & Weinstein, Y. *Phys. Rev. Lett.* **94**, 120501 (2005).

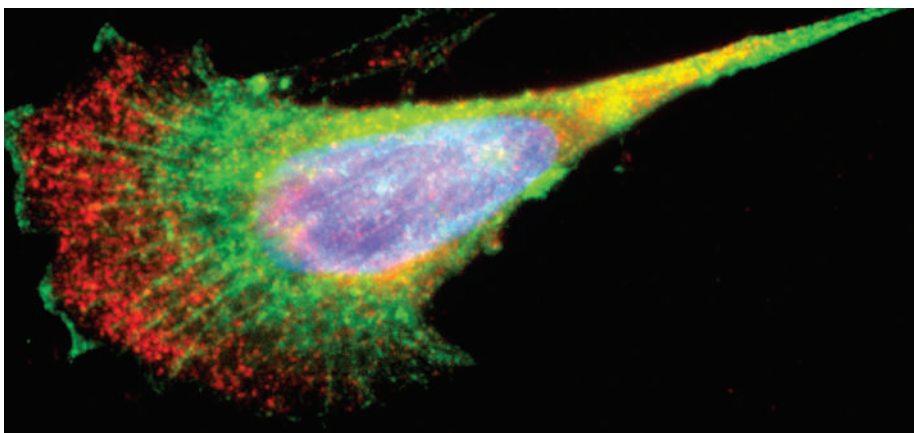


Figure 1 | A migrating fibroblast cell. β -actin mRNA (red) localizes to the fibroblast's leading edge. β -actin protein is shown as green, and the nucleus is stained blue.

A. WELLS

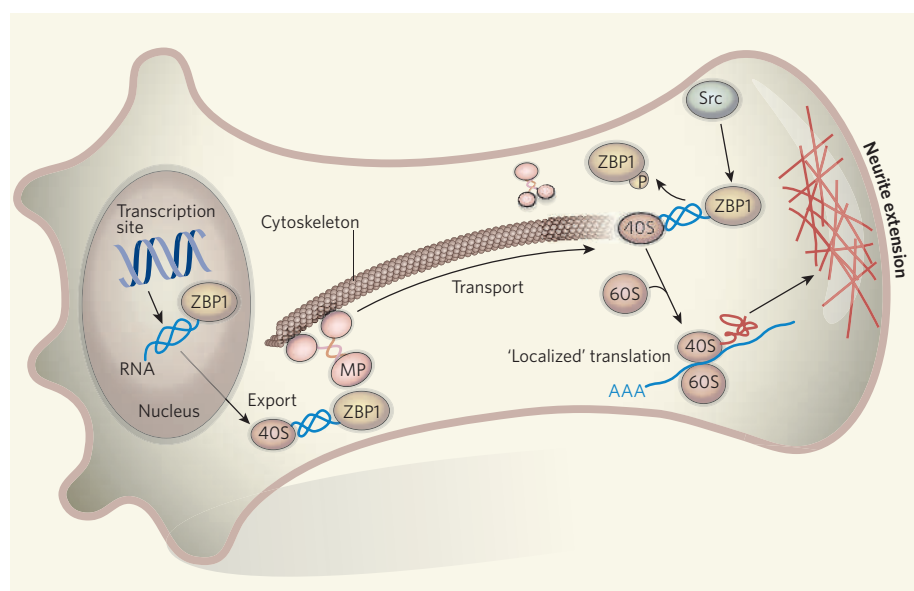


Figure 2 | Regulation of localized β -actin mRNA translation in a polarized neural cell. Hüttelmaier *et al.*⁴ propose that the ZBP1 protein controls the transport of β -actin mRNA and its subsequent translation into β -actin protein. ZBP1 associates with the β -actin mRNA in the nucleus and is exported to the cytoplasm. Here, the ZBP1- β -actin RNA complex binds to a motor protein (MP) and is transported along the cytoskeleton, the cell's internal scaffolding, to the periphery. During transport, ZBP1 prevents the mRNA from being translated into protein. When the ZBP1- β -actin RNA complex reaches its destination near the plasma membrane, ZBP1 is phosphorylated (P) by the non-receptor tyrosine kinase Src. This releases the mRNA, allowing the 40S and 60S subunits of the ribosomes to assemble and synthesize β -actin protein (red). The monomeric β -actin protein then assembles into the 'subcortical actin cytoskeleton', which pushes the leading edge onwards.

activity of other RNA-binding proteins⁶. So to investigate whether phosphorylation of ZBP1 modulates its regulatory role in translation, the authors used cells lacking wild-type ZBP1 and introduced into them a mutant ZBP1 that cannot be phosphorylated. This mutant ZBP1 could no longer repress translation of the β -actin mRNA, suggesting that phosphorylation by Src is crucial for translational regulation by ZBP1.

Where does this regulatory step occur? Hüttelmaier *et al.* next used fluorescence imaging to watch ZBP1 and Src in neuroblastoma cells. The two proteins came together only at the base of filopodia and in growth cones — motile, actin-rich structures that lead the way for outgrowing neurites. To obtain evidence that this interaction is functionally significant, the authors examined neurite outgrowth in cells lacking ZBP1. These cells have much shorter projections than usual, but adding ZBP1 back into the cells allowed them to grow normal-looking neurites. Adding the mutant, phosphorylation-incompetent ZBP1 did not produce normal outgrowths, and markedly reduced the amount of newly synthesized actin at the cell's periphery.

This is the first evidence that tyrosine phosphorylation of ZBP1 induces *de novo* synthesis of β -actin in a cellular compartment. It implies that ZBP1 could control a range of cellular processes, including cell migration and the formation of cellular polarity, especially the establishment of neuronal connections. The findings suggest a multi-step model for

the regulation of mRNA transport and translation (Fig. 2): in the nucleus, RNAs that will act at specific locations associate with corresponding RNA-binding proteins ('nuclear priming'). Once assembled, these 'transport-competent' RNA-protein complexes are exported into the cytoplasm⁷, where they associate with the cytoskeleton (the cell's internal scaffolding) and are transported to the cell's periphery with the help of molecular motors⁸. During their journey, the transcripts are translationally repressed by their protein partners, but on arrival at their destination they undergo a spatially controlled derepression to initiate translation in specific compartments of a cell.

This study generates many interesting questions that now need to be addressed. First, the mechanism of how ZBP1 controls translation is unknown, although the authors present preliminary evidence that unphosphorylated ZBP1 may inhibit the joining of the 40S and 60S subunits of the ribosome (the protein synthesis machinery). Second, it remains to be shown that the phosphorylation of ZBP1 regulates its RNA-binding capacity in an intact cell — there might be additional mechanisms that control ZBP1 function. Third, we do not understand how β -actin mRNA translation is restricted to the leading edge of a fibroblast or to the growth cones of developing neurons, rather than occurring all round the cell's periphery. Is Src kinase localized in a more restrictive manner than generally assumed, or is it spatially regulated? Fourth, how does this



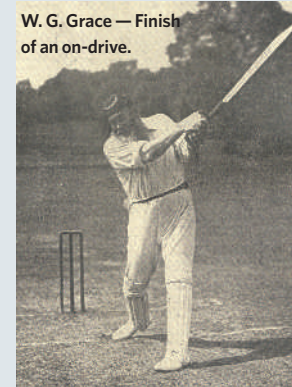
50 YEARS AGO

"Laboratory design" — It was decided to carry out a survey of the use actually made of space and services by scientists working in reasonably well-provided laboratories... Differences in the [bench] lengths used by scientific and experimental officers were small; it was found for these grades that about 12ft. of benching satisfied one man's requirements for 97 per cent of the time... A finding of some interest was that for 57 per cent of a scientist's time and 33 per cent of an assistant's time no bench was in use at all.

From *Nature* 26 November 1955.

100 YEARS AGO

Great Batsmen, their Methods at a Glance. By G. W. Beldam & C. B. Fry. Pp. xiv+716; illustrated by 600 Action photographs. Price 21s. net.



Each of the many batsmen pictured has been photographed in one or more characteristic attitudes before, during or after the striking of the ball, and after a careful study of every picture, Mr Fry has set down his own interpretation for the guidance of the reader... W. G. Grace, for example, is shown in twenty-six different attitudes, and all have some lesson to tell. In the photograph reproduced we have the finish of an on-drive, in which the turn of the body has aided powerfully in giving full effect to the stroke. The eyes are still looking at the spot where the ball was when it was struck. The whole series of photographs prove that all great batsmen follow the ball with their eye right up to the moment of striking.

From *Nature* 23 November 1905.

50 & 100 YEARS AGO

newly discovered function of ZBP1 contribute to biological and pathological processes involving cell migration? Such processes occur, for example, during embryonic development, infiltration of tissues by immune cells, wound healing and metastasis. Finally, does ZBP1 also act as a translational repressor in polarized neurons? If so, it will be exciting to discover whether it is involved in processes such as

neurite outgrowth, axon guidance and synaptic plasticity, which underlie learning and memory.

Ralf Dahm and Michael Kiebler are at the Center for Brain Research, Medical University of Vienna, Spitalgasse 4, 1090 Vienna, Austria.
e-mails: Ralf.Dahm@meduniwien.ac.at; Michael.Kiebler@meduniwien.ac.at

1. St Johnston, D. *Nature Rev. Mol. Cell Biol.* **6**, 363–375 (2005).
2. Huang, Y. S. & Richter, J. D. *Curr. Opin. Cell Biol.* **16**, 308–313 (2004).
3. Kloc, M., Zearfoss, N. R. & Etkin, L. D. *Cell* **108**, 533–544 (2002).
4. Hüttelmaier, S. *et al. Nature* **438**, 512–515 (2005).
5. Condeelis, J. & Singer, R. H. *Biol. Cell* **97**, 97–100 (2005).
6. Ostareck-Lederer, A. *et al. Mol. Cell Biol.* **22**, 4535–4543 (2002).
7. Kiebler, M. A., Jansen, R. P., Dahm, R. & Macchi, P. *Trends Biochem. Sci.* **30**, 228–231 (2005).
8. Jansen, R. P. *Nature Rev. Mol. Cell Biol.* **2**, 247–256 (2001).

CONDENSED-MATTER PHYSICS

Focus on the Fermi surface

Peter Littlewood and Šimon Kos

The electrical resistance of some manganese oxides takes a tumble when they become magnetic. Close examination confirms the interplay of conduction electrons and lattice vibrations that contributes to this effect.

Certain manganese oxides (manganites) exhibit an intriguing effect known as ‘colossal magnetoresistance’. Below a certain critical temperature, these materials become ferromagnetic — showing the spontaneous alignment of electron spins that accounts for the magnetic attraction of materials such as iron — and this is accompanied by a drastic reduction in electrical resistance. On page 474 of this issue, Mannella and colleagues¹ describe the electronic properties of a two-layer manganite compound, $\text{La}_{1.2}\text{Sr}_{1.8}\text{Mn}_2\text{O}_7$ (LSMO), revealing that the interaction of electrons and lattice vibrations known as phonons is crucial to colossal magnetoresistance. The results also bring to light unexpected similarities between the electronic structures of manganites and superconducting copper oxides.

In a metallic material such as a manganite, electrons fill quantum-mechanically allowed energy states singly from the lowest possible energy upwards. (This is a consequence of the Pauli exclusion principle, which holds that no two electrons may share the same quantum state.) The energy of the most highly occupied state is known as the chemical potential, or Fermi energy. The electronic energy states in a solid with a periodic lattice structure also have a well-defined momentum; when components of these momenta in the three spatial dimensions are plotted against each other, the occupied states form a characteristic shape bounded by a so-called Fermi surface. Only electrons in states near the Fermi surface — those with the highest momenta — contribute to conduction. According to a central tenet of quantum mechanics, known as Heisenberg’s uncertainty relation, how sharply defined these energy states are is a measure of the degree to which electrons scatter on the lattice or on each other. So a good metal, with a low electrical resistance (little scattering), will have a sharply delineated Fermi surface.

Mannella and colleagues¹ provide the first experimental observation of the Fermi

surface in a manganite. The technique they use, photoemission spectroscopy, is based on the photoelectric effect, in which light striking a metallic surface behaves as if it were a particle, knocking out a loosely bound electron. The electron’s momentum and energy can be probed directly using this method, and the photoelectric effect has been a workhorse of experimental condensed-matter physics since it was first explained by Albert Einstein 100 years ago (for which he won the Nobel Prize in Physics in 1921).

The authors show that the Fermi surface of LSMO becomes more sharply defined when the material is cooled into the ferromagnetic state, indicating that its resistance has fallen. The result fits in with our understanding of colossal magnetoresistance as the suppression, induced by the onset of ferromagnetic order, of an interaction between electrons and phonons (the quanta of lattice vibrations)² that increases resistance. This bundling of electron and lattice properties can itself be treated as a physical entity moving through the lattice — a ‘quasiparticle’ known as a polaron. Scanning tunnelling microscopy measurements of LSMO support this picture³, showing images of polarons trapped by occasional impurities.

Mannella and colleagues’ results also indicate that the spectral weight of the sample (loosely, the proportion of the total number of energy states that exist at the Fermi surface) is very small, explaining why these states have not been observed previously. In addition, the measured energy spectrum at the Fermi surface is not isotropic, but depends strongly on the direction: electrons propagate readily (albeit with a velocity five times smaller than expected) in a direction that is diagonal to the square lattice of manganese atoms, but poorly along the axes of the lattice.

The reduced spectral weight and velocity seem to imply that, even in the metallic state, in which conduction electrons supposedly move freely throughout the lattice,

electrons and phonons are interdependent. The results are puzzling, because at low temperatures manganites such as LSMO are good metals with an isotropic conductivity. The shape of the Fermi surface itself — with large parts that are nearly parallel, or ‘nested’ — supplies one possible interpretation. Nesting provides a channel through which an electron can be scattered between different parts of the Fermi surface; in this case it is scattered by a phonon, but in general it could also be scattered by magnetic fluctuations, should these exist. Scattering would reduce the electronic spectral weight around the Fermi energy and induce a gap in the energy spectrum.

The spectrum measured by Mannella *et al.* is very similar to that of the mysterious ‘pseudogap’ phase seen in high-temperature cuprate superconductors. This intriguing fact suggests that such a gap is a generic feature of the oxides of transition metals — rather than being a facet only of cuprate superconductors, as had been assumed. A further widespread belief is that phonons play no role in the high-temperature superconductivity seen in cuprates, despite the fact that interactions between electrons and phonons underlie conventional, low-temperature superconductivity. But it has been shown^{4–6} that phonons affect various properties of the electrons in these superconductors, especially in the pseudogap phase. Mannella *et al.*¹ provide yet another incentive to examine the role of phonons more carefully.

Peter Littlewood and Šimon Kos are at the Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK.
e-mail: pbl21@cam.ac.uk

1. Mannella, N. *et al. Nature* **438**, 474–478 (2005).
2. Millis, A. J. *Nature* **392**, 147–150 (1998).
3. Ronnow, H. M., Renner, Ch., Aeppli, G., Kimura, T. & Tokura, Y. preprint available at www.arxiv.org/cond-mat/0511138 (2005).
4. Lanzara, A. *et al. Nature* **412**, 510–514 (2001).
5. Gweon, G.-H. *et al. Nature* **430**, 187–190 (2004).
6. Dordevic, S. V. *et al. Phys. Rev. B* **71**, 104529 (2005).

PHOTONICS

Winged light

In man-made optical emitters such as LEDs, paradoxically most of the light remains trapped inside. So physicists have recently begun to exploit structures known as photonic crystals to try to extract more light. But while they have been trying to forge the most efficient devices, it now turns out that the swallowtail butterfly has already mastered the art of emitting light.

The eye-catching *Papilio nireus* butterfly from eastern and central Asia has wings emblazoned with iridescent blue-green patches (pictured). Pete Vukusic and Ian Hooper have used electron microscopy to learn more about the optical properties of these spots (*Science* **310**, 1151; 2005). Their

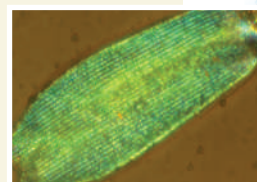
images reveal that the scales on the butterfly's wings contain an intricate nanostructure of all-natural photonic crystals.

The crystals are just one part of a sophisticated photonic system that provides intense, directed light and endows the butterfly with its striking colours. A two-dimensional photonic crystal 'slab', about two micrometres thick, is set within the solid outer wing layer, and consists of hollow cylinders arranged in triangular domains. Highly fluorescent pigment, whose peak excitation wavelength matches that of the blue of the sky, is infused throughout the slab. And layered structures, known as Bragg reflectors, sit about 1.5 micrometres below the crystals.

Photonic crystals can stop light of specific frequencies from propagating within them. In this way, they can be used to limit or inhibit the emission of an optical emitter that is embedded deep within the crystal.

The optical properties of the crystals in this butterfly's scales are such that emission from the fluorescent dye is inhibited only within the crystal plane. Added to the effects of the underlying Bragg reflectors, this means the fluorescent light has nowhere to

go but up and out. The effect is clearly seen in the optical-microscope image of a single scale (inset). Notably, similar techniques are used in the latest ultra-efficient LEDs.



An astonishing variety of natural photonic structures are being uncovered not just in butterflies,

but also in other insects, birds and fish. Although nature and technology have evolved independently, they have ultimately come up with the same design.

Amber Jenkins

P. VUKUSIC, UNIV. EXETER

OBESITY

Aquaporin enters the picture

Gema Frühbeck

The aquaporins are membrane channels that were originally identified as regulators of a cell's water balance. A member of the aquaporin family is now implicated as a central agent in controlling fat metabolism.

Mammals, like other organisms, have had to cope with feast and famine over evolutionary time. Hence the selection of genes that in times of plenty allow fat accumulation — principally in the form of triglycerides in fat cells — and in times of need promote the mobilization of those reserves. But the complex pathways involved in controlling this energy balance, which are implicated in the development of obesity and diabetes, are far from fully understood.

Hibuse *et al.*¹, writing in *Proceedings of the National Academy of Sciences*, add to the picture by showing that deficiency of a certain integral membrane protein in fat cells, called aquaporin-7, is associated with adult-onset obesity in mice. Together with previous work^{2,3}, the finding provides evidence not only that aquaporin-7 transports glycerol *in vivo*, but also that fat-cell permeability to glycerol is a key element in regulating fat accumulation.

The aquaporins are a family of proteins that facilitate the movement of water across the cell membrane, and in some cases they also act as a channel for molecules such as glycerol and other small solutes⁴. Aquaporin-7 is one such channel⁵⁻⁷, and its synthesis in fat cells

has been reported to be sensitive to various stimuli, including fasting and insulin.

Triglycerides consist of three fatty acids bound to a glycerol backbone. They are stored in fat tissue to serve as the body's principal fuel reserve. Fatty acids are mobilized to provide energy during prolonged exercise or starvation, and are used preferentially as fuel by muscle and liver. Glycerol, meanwhile, is both a triglyceride precursor, as glycerol-3-phosphate, and along with free fatty acids is one of the initial metabolic products of triglyceride mobilization. Following its release into the bloodstream from fat cells, glycerol is taken up mainly by the liver and converted to glucose. During fasting, glucose output from the liver is the main source of blood glucose, and the release of glycerol from fat cells is a major source of the necessary substrate. Thus, during fasting, effective systems are required to increase glycerol release from fat cells and its uptake by the liver. Loss of glycerol transport affects the levels of metabolites and hormones in the blood, and interferes with the triglyceride-fatty-acid cycle, as well as with glucose metabolism in the liver and in muscle (a glucose consumer).

The latest insights into aquaporin-7 function have largely stemmed from two independently created¹⁻³ lines of mice that lack the protein. In one set of findings^{1,2}, these knockout mice maintained similar body weights to normal mice until 12 weeks of age, at which time they started to accumulate fat and become heavier, even though their food intake was the same as that of normal mice. The other line of knockout mice³ did not put on excess weight, a difference that Hibuse *et al.*¹ ascribe to the different genetic backgrounds of the two lines. But in other respects, such as the remarkable fat-cell enlargement seen, the mice showed similar responses to aquaporin deletion.

The enlarged fat cells resulted from the larger size of lipid droplets. Lipid synthesis and breakdown were unaffected by deletion of aquaporin-7, but there was a threefold decrease in glycerol release³. Thus, the reduced membrane permeability to glycerol results in an increase in steady-state glycerol concentration, which leads to increased glycerol-3-phosphate and hence triglyceride biosynthesis (Fig. 1). Activation of glycerol kinase, the enzyme that catalyses the conversion of glycerol to glycerol-3-phosphate, favours recycling of free fatty acids and results in a progressive accumulation of triglycerides¹⁻³.

In the latest work, Hibuse *et al.*¹ observed that 20-week-old obese knockout mice showed severely impaired glucose tolerance, seen as abnormally high blood-glucose levels. This was accompanied by decreased effects of insulin on fat tissue, liver and muscle. Such signs are characteristic of the whole-body insulin resistance associated with adult-onset obesity. Under normal circumstances, insulin

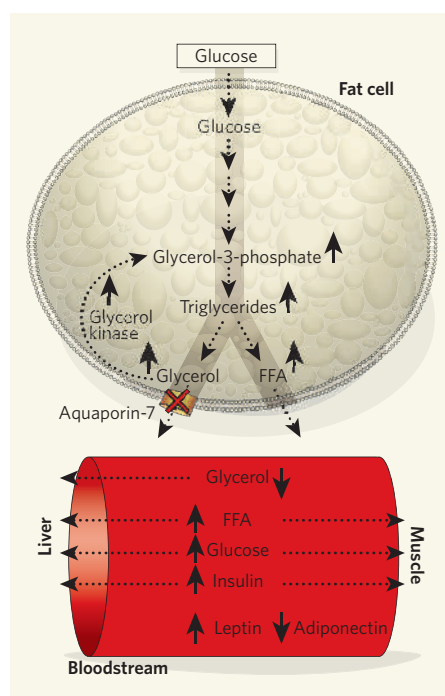


Figure 1 | Consequences of aquaporin-7 deficiency¹⁻³. In fat cells, glucose is the substrate for the synthesis of fat deposits in the form of triglycerides, which can subsequently be mobilized as glycerol and free fatty acids (FFA). If the gene for aquaporin-7 is deleted, fat cells show glycerol accumulation, which results in enhanced glycerol kinase activity and triglyceride synthesis, and fat-cell enlargement. The loss of aquaporin-7-mediated glycerol transport from fat cells affects the concentrations of various metabolites and hormones in the bloodstream, as shown. The upshot is perturbation of the triglyceride–fatty-acid cycle, and of glucose metabolism in the liver and muscle. Dotted arrows indicate pathways; solid arrows indicate increases or decreases in concentrations.

represses expression of the aquaporin-7 gene^{6,7}. In the insulin-resistant condition, however, the sustained increase of aquaporin-7 production in fat tissue perturbs glucose balance by producing an abnormal increase in glucose production in the liver.

Further characteristics of the older knockout mice were low bloodstream levels of glycerol and of adiponectin, a secretion product of fat cells that mediates insulin action (Fig. 1). Conversely, these mice had high blood concentrations of glucose, insulin, free fatty acids and leptin (leptin participates in body-weight regulation through its effects on the hypothalamus, the brain region that controls feeding behaviour, body temperature and energy output). Hibuse *et al.* found that these metabolic perturbations were more severe in aquaporin-deficient mice fed a high-fat diet. In tying these various observations together, there is evidently a need to invoke the existence of a coordinated regulation of fat-cell-specific and liver-cell-specific glycerol channels in determining glucose metabolism and insulin resistance⁸.

Ten-week-old knockout mice, by contrast, did not show these various symptoms¹. They

were not insulin resistant, and had low levels of glucose in the bloodstream. This was probably due to reduced glycerol transport to the liver and inefficient glucose production there — despite the increased expression of phosphoenolpyruvate carboxykinase, the key liver enzyme in the pathway that generates glucose from non-carbohydrate substrates such as glycerol⁹.

All in all, these three studies¹⁻³ provide firm evidence that aquaporin-7 is required for glycerol release from fat cells in mice. The latest work¹ also shows that this channel is a broader player in glucose balance and sensitivity to insulin. But how relevant are the findings for human obesity?

The human gene that encodes aquaporin-7 lies on the long arm of chromosome 13, and has not previously been implicated in obesity. So far, only a single human case of loss-of-function mutation in the aquaporin-7 gene has been reported⁷: the subject did not become obese or diabetic, the only apparent consequence being impaired increase of glycerol in the bloodstream during exercise.

Nonetheless, it might be worth exploring the possibility of selectively enhancing aquaporin-

7 expression in fat tissue for therapeutic ends. Another obvious project would be to look for differences in aquaporin-7 production and function between lean and obese individuals. Other steps would be to investigate how the molecular structure of the pore affects its specificity for glycerol, with drug design in mind, and to carry out gene-expression and protein-stability studies to achieve a better basic understanding of aquaporin action.

Gema Frühbeck is in the Department of Endocrinology, Clínica Universitaria, Universidad de Navarra, 31008 Pamplona, Spain. e-mail: gfruhbeck@unav.es

- Hibuse, T. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 10993–10998 (2005).
- Maeda, N. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 17801–17806 (2004).
- Hara-Chikuma, M. *et al.* *J. Biol. Chem.* **280**, 15493–15496 (2005).
- King, L. S., Kozono, D. & Agre, P. *Nature Rev. Mol. Cell Biol.* **5**, 687–698 (2004).
- Kuriyama, H. *et al.* *Biochem. Biophys. Res. Commun.* **241**, 53–58 (1997).
- Kishida, K. *et al.* *J. Biol. Chem.* **276**, 36251–36260 (2001).
- Kondo, H. *et al.* *Eur. J. Biochem.* **269**, 1814–1826 (2002).
- Kuriyama, H. *et al.* *Diabetes* **51**, 2915–2921 (2002).
- MacDougald, O. A. & Burant, C. F. *Proc. Natl Acad. Sci. USA* **102**, 10759–10760 (2005).

HISTOCOMPATIBILITY

Colonial match and mismatch

Gary W. Litman

Distinguishing self from non-self is the underlying basis of immunity. Intriguingly, the genetic system that governs a natural process akin to tissue transplantation in vertebrates has been characterized in an invertebrate.

On page 454 of this issue, a remarkable paper by De Tomaso and colleagues¹ describes discoveries on two different fronts. The authors' subject of study is a small marine organism, a tunicate called *Botryllus schlosseri* (Fig. 1, overleaf). Their results provide insight into why colonies of this animal sometimes merge into one, with a common blood circulation, and sometimes reject that option. Also — and more notably — the results will raise awareness of forms of immune recognition that extend beyond those known in vertebrates.

The ability of an individual to distinguish tissues of another individual from its own is a fundamental characteristic of multicellular animals². In humans and other vertebrates, this process, known as allorecognition, is seen in the success or failure of skin grafts and organ or bone-marrow transplants. These interactions are determined by the products of a particularly diverse group of genes encoded within the major histocompatibility complex, known as MHC-I and -II, which primarily mobilize the immune response to foreign bodies, including infections by viruses and bacteria. It is their polymorphism — the extensive

differences between the products of individual genes — and their expression in tissues throughout the body that make these particular MHC products such central players in vertebrate histocompatibility reactions.

In the animal world, only jawed vertebrates possess MHC-I and -II. Naturally occurring allogeneic interactions have, however, been well documented in protochordates and the non-chordate invertebrates (see Box 1 for further explanation). In *Botryllus*, allorecognition is governed at a single genetic locus, FuHC (for fusibility/histocompatibility). De Tomaso *et al.*¹ have now isolated the FuHC gene locus and provide the first insights into the molecular basis for this ancient form of MHC-independent allorecognition.

The fusibility seen in *Botryllus* involves ampullae, small protrusions at the periphery of tunicate colonies. When the ampullae from two colonies come into close proximity, the products of the FuHC gene locus determine whether fusion or rejection occurs. Fusion results when two colonies share at least one FuHC gene variant, or allele; rejection occurs when no FuHC alleles are shared. Thus, for

example, a colony with AA alleles at the FuHC locus would fuse with one with AB, but AA × BB would result in rejection.

Most populations of *Botryllus* consist of many genetically disparate colonies, and tens to hundreds of allelic variants of FuHC probably exist. Individual colonies are preferentially maintained as heterozygotes — that is, with two different alleles at the FuHC locus — because of the developmentally regulated advantage conferred through heterozygosity^{3,4}.

De Tomaso *et al.*¹ isolated and physically mapped the *Botryllus* FuHC locus by classical genetic approaches. Its gene product would be expected to span the cell membrane and be highly polymorphic, as turned out to be the case; further study showed that it contains units called epidermal growth factor repeats and also immunoglobulin domains, but of a different type from those seen in MHC-I and -II. The authors also found that allelic polymorphism of FuHC was correlated with fusion or rejection in transplantation assays of laboratory strains. They went further by observing the interaction of wild-type colonies harvested from different geographical areas, and the results confirmed the absolute concordance of fusibility and genotype.

Botryllus colonies live in close proximity, in the narrow ecological niches of tidal pools; colony mergers, with the transfer of genetic material, are an essential part of their population dynamics. The mechanistic basis for this has long been thought to involve genetically determined stem-cell parasitism⁵, in which cells mobilize and transfer between individuals, potentially being able to replace the cells of the host. The characteristic features of the parasitic



ANDREW N. COHEN

Figure 1 | *Botryllus schlosseri*: test case in histocompatibility. The *Botryllus* life cycle includes both a free-living sexual stage and a sedentary, asexual stage in which colonies grow by budding. A colony (pictured here) consists of genetically identical individuals known as zooids, which are sheathed in a gelatinous tunic and are connected by a vascular system. When colonies come into close contact, histocompatibility reactions occur at their peripheries, in small protrusions termed ampullae. Colonies either fuse (which leads to sharing a vascular system) or reject fusion on the basis of allele compatibility or incompatibility, respectively, at the FuHC locus. See also Fig. 1 on page 455 of the paper¹.

cells are retained upon surgical transplantation (D. J. Laird and A. W. De Tomaso, personal communication). The FuHC genes have probably evolved to restrict germline exchange (exchange of genetic material that will be passed on to the next generation) to closely related individuals as a way of preserving a diverse gene pool that otherwise could become homogenized throughout an entire population. This, then, is a fascinating case of potentially rapid mechanisms of germline change that is separate from traditional inheritance. Similar effects probably account for polymorphic histocompatibility in other invertebrates⁶.

The histocompatibility system revealed by De Tomaso *et al.* is evidently essential to the survival of *Botryllus*. Given its importance, does a similar system survive in higher vertebrates? In their paper, the authors draw analogies between fusibility and recognition of 'missing self', in which the lack of at least one interaction between FuHC and its putative receptor triggers a rejection response. In mammals, and probably other higher vertebrates, recognition of missing self is the basis for immunity mediated by natural killer cells; these typically sense either a decrease in polymorphic MHC-I (a consequence of viral infection) or an increase in stress-induced MHC-I-related determinants⁷, both of which seem to be absent in another protochordate studied⁸. In this and other work that delves deep into evolutionary history, the overriding issue is whether the evolutionary trail is clear enough to allow recognition of traditional relationships (such as genetic sequence identity, protein structure homologies or

regulatory network similarities). De Tomaso *et al.*¹ have identified polymorphic genes that map to within 200 kilobases of the FuHC locus and that could encode a receptor for the FuHC gene product. This finding should clarify the relatedness of fusibility to other self–non-self recognition processes.

Irrespective of whether this ancient adaptation may live on elsewhere, colonial fusibility is the best-characterized example of the mechanisms that invertebrates, protochordates and jawless vertebrates use to distinguish self from non-self. Over the past several years, our eyes have been opened to a variety of alternative mediators of immune function that have blurred the traditional distinctions between innate and adaptive immunity⁹. This latest paper takes the story further, and is probably destined to become a classic. It reveals that genetically driven allorecognition may have been essential in creating and maintaining the fittest gene pools in our ancient ancestors. ■

Gary W. Litman is at the Children's Research Institute, 830 First Street South, St Petersburg, Florida 33701, USA.
e-mail: litmanG@allkids.org

Box 1 Chordates and invertebrates

The following is a summary of the relationships among the animal groups mentioned in this article. The Chordata include the Vertebrata, Cephalochordata and Urochordata, of which *Botryllus schlosseri* is a member. The Vertebrata include both jawed and jawless animals, but the only surviving members of the latter are the lampreys and hagfishes. The Protochordata are an informal grouping of the Cephalochordata and Urochordata, and another group, the Hemichordata. As invertebrate chordates, the Protochordata are not to be confused with the familiar non-chordate invertebrates (echinoderms, molluscs, arthropods and so on).

Traditionally, both the invertebrates and protochordates have been considered to possess innate immunity, which is present from birth and is not highly specific; general inflammation is an example of an innate immune response. Vertebrates possess both innate and adaptive immunity; the latter is triggered in response to a foreign challenge, such as an infection, and is highly specific. In jawed vertebrates, rejection of a foreign tissue graft is mediated by the cellular arm of the adaptive immune system through interactions with products of the major histocompatibility complex.

G.W.L.

1. De Tomaso, A. W. *et al.* *Nature* **438**, 454–459 (2005).
2. Burnet, F. M. *Nature* **232**, 230–235 (1971).
3. Scofield, V. L., Schlumpberger, J. M., West, L. A. & Weissman, I. L. *Nature* **295**, 499–502 (1982).
4. De Tomaso, A. W. & Weissman, I. L. *Science* **303**, 977 (2004).
5. Stoner, D. S., Rinkevich, B. & Weissman, I. L. *Proc. Natl Acad. Sci. USA* **96**, 9148–9153 (1999).
6. Buss, L. W. *The Evolution of Individuality* (Princeton Univ. Press, 1987).
7. Vivier, E. & Malissen, B. *Nature Immunol.* **6**, 17–21 (2005).
8. Dehal, P. *et al.* *Science* **298**, 2157–2160 (2002).
9. Litman, G. W., Cannon, J. P. & Dishaw, L. J. *Nature Rev. Immunol.* **5**, 866–879 (2005).

BRIEF COMMUNICATIONS

Engineering *Escherichia coli* to see light

These smart bacteria 'photograph' a light pattern as a high-definition chemical image.

We have designed a bacterial system that is switched between different states by red light. The system consists of a synthetic sensor kinase that allows a lawn of bacteria to function as a biological film, such that the projection of a pattern of light on to the bacteria produces a high-definition (about 100 megapixels per square inch), two-dimensional chemical image. This spatial control of bacterial gene expression could be used to 'print' complex biological materials, for example, and to investigate signalling pathways through precise spatial and temporal control of their phosphorylation steps.

Plants and some bacteria use a class of protein photoreceptors known as phytochromes to control phototaxis, photosynthesis and the production of protective pigments^{1–3}. Photoreceptors are not found in enterobacteria, such as *Escherichia coli*, so we created a light sensor that functions in *E. coli* by engineering a chimaera that uses a phytochrome from a cyanobacterium.

A phytochrome is a two-component system that consists of a membrane-bound, extra-cellular sensor that responds to light and an intracellular response-regulator¹. The response-regulators of most phytochromes do not have DNA-binding domains and do not directly regulate gene expression, so we fused a cyanobacterial photoreceptor to an *E. coli* intracellular histidine kinase domain (Fig. 1a, and see supplementary information). This design was based on the well studied *E. coli* EnvZ–OmpR two-component system, which normally regulates porin expression in response to osmotic shock⁴. The EnvZ histidine kinase domain has been used for the construction of functional chimaeras^{5,6}, and a plant phytochrome has previously been used to construct a two-hybrid gene expression system in yeast⁷.

To create the chimaera, we aligned members of the phytochrome family with EnvZ and identified potential functional crossover points between the *Synechocystis* phytochrome Cph1 and EnvZ. (For methods, see supplementary information.) The length and composition of the peptide that links a photoreceptor to its response-regulator can affect signal transduction^{5,6}, and we therefore constructed a series of chimaeras with variable linker lengths. The variants were transformed into a Δ EnvZ *E. coli* strain containing a chromosomal fusion between the OmpR-dependent *ompC* promoter and the *lacZ* reporter⁴, which

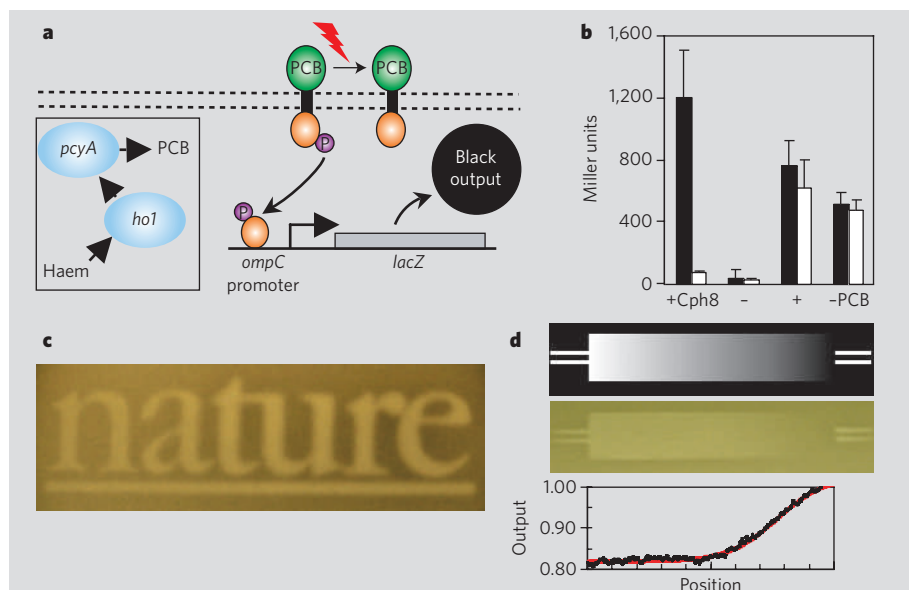


Figure 1 | Light imaging by engineered *Escherichia coli*. **a**, The chimaeric light receptor Cph8 contains the photoreceptor from Cph1 (green) and the histidine kinase and response-regulator from EnvZ–OmpR (orange); inset, conversion of haem to phycocyanobilin (PCB), which forms part of the photoreceptor. Red light drives the sensor to a state in which autophosphorylation is inhibited (right), turning off gene expression. For details of genes, see text. **b**, Miller assay showing that Cph8 is active in the dark (black bars) in the presence of PCB and inactive in the light (white bars). There is no light-dependent activity in the absence of Cph8 (–) and there is constitutive activity when only the histidine kinase domain of EnvZ is expressed (+), or when the PCB metabolic pathway is not included (– PCB). **c**, When an image is projected on to a bacterial lawn, the *lacZ* reporter is expressed only in the dark regions. **d**, Transfer function of the circuit. As the intensity of the light is increased by using a light gradient projected from a 35-mm slide, the circuit output gives a graded response.

enzymatically produces a black compound.

The part of the photoreceptor that responds to light, phycocyanobilin, is not naturally produced in *E. coli*. We therefore introduced two phycocyanobilin-biosynthesis genes (*ho1* and *pcyA*) from *Synechocystis* that convert haem into phycocyanobilin⁸ (parts BBa_I15008, BBa_I15009; MIT Registry of Standard Biological Parts) (Fig. 1a, inset). Individual Cph1–EnvZ chimaeras were then activated at 37 °C for 4 h with broad-spectrum light and assayed for expression of the *lacZ* reporter. The chimaera Cph8 (BBa_I15010) produced a particularly strong response to light (Fig. 1b).

For bacterial photography, we grew a lawn of bacteria on agar. The *lacZ* reporter was visualized by addition of S-gal (3,4-cyclohexenoesculetin- β -D-galactopyranoside): LacZ catalyses the formation of a stable, insoluble, black precipitate from S-gal. Light repressed gene expression in the bacteria, giving a high-contrast replica of the applied image on

the biological film, in which light regions appeared light and dark regions were dark (Fig. 1c, and see supplementary information). The *lacZ* activity showed a graded response to increasing light intensity that was minimal in the brightest light (Fig. 1d).

Our creation of a novel genetic circuit with an image-processing function demonstrates the power and accessibility of the tool sets and methods available in the nascent field of synthetic biology. The principle of programmed light regulation should enable gene expression to be spatially and temporally controlled in individual cells and in populations, leading to potential application in bacterial microlithography, manufacture of biological material composites and the study of multicellular signalling networks.

Anselm Levskaya*, Aaron A. Chevalier†, Jeffrey J. Tabor†, Zachary Booth Simpson†, Laura A. Lavery†, Matthew Levy†, Eric A. Davidson†, Alexander Scouras†, Andrew D. Ellington†‡, Edward M. Marcotte†‡, Christopher A. Voigt*§||

*Biophysics Program, University of California, San Francisco, California 94143, USA

e-mail: cavoigt@picasso.ucsf.edu

†Center for Systems and Synthetic Biology and Institute for Cell and Molecular Biology, and ‡Department of Chemistry and Biochemistry, University of Texas, Austin, Texas 78712, USA

§Department of Synthetic Biology, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

||Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94107, USA

1. Yeh, K.-C., Wu, S.-H., Murphy, J. T. & Lagarias, J. C. *Science* **277**, 1505–1508 (1997).
2. Schmitz, O., Katayama, M., Williams, S. B., Kondo, T. & Golden, S. S. *Science* **289**, 765–768 (2000).
3. Davis, S. J., Vener, A. V. & Vierstra, R. D. *Science* **286**, 2517–2520 (1999).
4. Utsumi, R. *et al.* *Science* **245**, 1246–1249 (1989).
5. Jin, T. & Inouye, M. *J. Mol. Biol.* **244**, 477–481 (1994).
6. Kwon, O., Georgellis, D. & Lin, E. C. C. *J. Biol. Chem.* **278**, 13192–13195 (2003).
7. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. *Nature Biotechnol.* **20**, 1041–1044 (2002).
8. Gambetta, G. A. & Lagarias, J. C. *Proc. Natl Acad. Sci. USA* **98**, 10566–10571 (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/nature04405

INSECT COMMUNICATION

'No entry' signal in ant foraging

Forager ants lay attractive trail pheromones to guide nestmates to food^{1,2}, but the effectiveness of foraging networks might be improved if pheromones could also be used to repel foragers from unrewarding routes^{3,4}. Here we present empirical evidence for such a negative trail pheromone, deployed by Pharaoh's ants (*Monomorium pharaonis*) as a 'no entry' signal to mark an unrewarding foraging path. This finding constitutes another example of the sophisticated control mechanisms used in self-organized ant colonies.

To investigate whether foragers lay a negative signal on the unrewarding branch of a trail bifurcation, we removed paper substrate from immediately after the fork on the unrewarding branch (the other branch led to a sucrose feeder) after it had been used by a trail-laying colony of ants. This paper substrate was transferred to the entrance of one branch of a similar set-up, in which both branches had previously led to sucrose and had been used by a second colony of ants. The other branch of the second set-up received a neutral control paper substrate (for details, see supplementary information). Foragers walking from the nest could choose either of the test branches or make a U-turn.

We found that 69% continued to walk away from the nest and make a branch choice. Of these, most (71%) chose the branch with the control substrate ($\chi^2=22.1$, d.f.=1, $n=137$, $P<0.001$); the remainder U-turned towards the nest on reaching the trail bifurcation. U-turns were more than four times as likely if the ant had contacted the unrewarding-branch substrate (55%) as opposed to the neutral-control substrate (13%) ($\chi^2=40.9$, d.f.=1, $n=200$, $P<0.0001$). Neither substrate came from a previously rewarding trail, so this result cannot be attributed to differences in positive-trail pheromone concentrations.

We next investigated the negative signal's location by taking substrate from five locations on a bifurcating trail that had one rewarding

and one unrewarding branch. These sections, along with neutral controls, were tested on unbranched foraging trails (see supplementary information) by noting whether individual foragers walking over them did a U-turn. Compared with ants on the control substrate, almost twice as many ants U-turned when walking on substrate from the unrewarding branch near the bifurcation (N_b) (19% and 34%, respectively; $P<0.001$) (Fig. 1a). However, U-turns were as frequent on substrate from the unrewarding branch end (N_e) (27%) as on the control (27%) (NS) (Fig. 1a). Ants U-turned less often on sections from the rewarding trail (stem S, 12%; feeder branch close to the bifurcation F_b , 12%; and feeder-branch end F_e , 13%). These values are significantly lower than those for the relevant control (S, $P<0.001$; F_b , $P<0.05$; F_e , $P<0.001$) (Fig. 1a).

In the same experiment, we also determined whether foragers could detect the negative signal before reaching the substrate on which it had been laid, using walking behaviour (zigzagging versus walking straight) as a bio-assay. Our results show that significantly more ants zigzagged when approaching substrate from an unrewarding branch just after the bifurcation ($P<0.01$) or at the branch end ($P<0.05$) than did controls (Fig. 1b). Conversely, significantly fewer zigzagged when approaching substrate leading to the feeder (S, $P<0.01$; F_b , $P<0.05$; F_e , $P<0.05$) (Fig. 1b).

Our results show that Pharaoh's ants use a sophisticated trail system with a negative, repellent pheromone to mark unrewarding branches. The signal is concentrated at decision points — trail bifurcations⁵. As it is volatile, it provides advance warning — like human road signs situated before junctions. Across a trail network, the pheromone could help direct foragers to food by closing off unrewarding sections. Exactly how negative pheromones enhance foraging efficiency in trail networks is not known, but they might complement attractive trail pheromones^{6,7}

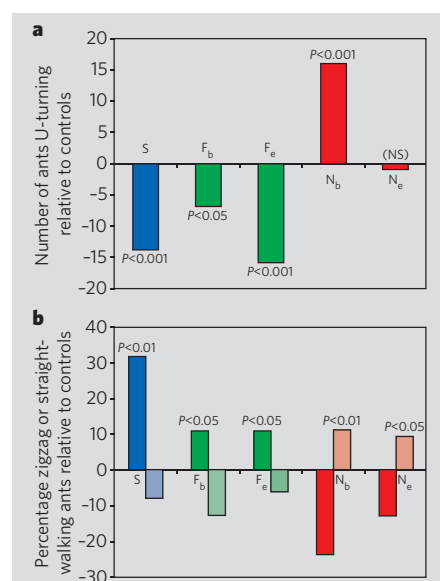


Figure 1 | Identifying the location of the negative pheromone. The ants' response is monitored by their walking behaviour, with U-turning or zigzagging on unbranched trails indicating detection. Test sections: S, 1 mm before bifurcation; F_b and N_b , 3 mm after bifurcation on feeder and non-feeder branches, respectively; F_e and N_e , 60 mm from bifurcation at the ends of feeder and non-feeder branches, respectively. (For details and chi-squared tests, see supplementary information.) **a**, Number of ants that U-turned while walking on different test sections, relative to controls. **b**, Percentage of straight-walking (left bars) or zigzagging (right bars) ants, relative to controls.

used by Pharaoh's ants in trail choice, or they could prevent strong positive feedback by attractive pheromones from locking the system into suboptimal solutions^{1,8}.

Elva J. H. Robinson*, Duncan E. Jackson†, Mike Holcombe†, Francis L. W. Ratnieks*

*Laboratory of Apiculture and Social Insects, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK
e-mail: e.robinson@dcs.sheffield.ac.uk

†Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK

1. Sumpter, D. J. T. & Beekman, M. *Anim. Behav.* **66**, 273–280 (2003).
2. Camazine, S. *et al.* *Self-Organization in Biological Systems* (Princeton Univ. Press, Princeton and Oxford, 2001).
3. Stickland, T. R., Britton, N. F. & Franks, N. R. in *Information Processing in Social Insects* (eds Detrain, C., Deneubourg, J. L. & Pasteels, J. M.) 83–100 (Birkhäuser, Basel, 1999).
4. Britton, N. F., Stickland, T. R. & Franks, N. R. *J. Biol. Syst.* **6**, 315–336 (1998).
5. Jackson, D. E., Holcombe, M. & Ratnieks, F. L. W. *Nature* **432**, 907–909 (2004).
6. Sudd, J. H. *Br. J. Anim. Behav.* **5**, 104–109 (1957).
7. Jeanson, R., Deneubourg, J. L. & Ratnieks, F. L. W. *Physiol. Entomol.* **28**, 192–198 (2003).
8. Beckers, R., Deneubourg, J. L., Goss, S. & Pasteels, J. M. *Insectes Soc.* **37**, 258–267 (1990).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/438442a

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

IMMUNOLOGY

Insulin auto-antigenicity in type 1 diabetes

Arising from: M. Nakayama *et al.* *Nature* **435**, 220–223 (2005) and S. C. Kent *et al.* *Nature* **435**, 224–228 (2005)

Spontaneous type 1 diabetes occurs when the autoimmune destruction of pancreatic β -islet cells prevents production of the hormone insulin. This causes an inability to regulate glucose metabolism, which results in dangerously raised blood glucose concentrations. It is generally accepted that thymus-derived lymphocytes (T cells) are critically involved in the onset and progression of type 1 diabetes, but the antigens that initiate and drive this destructive process remain poorly characterized — although several candidates have been considered¹. Nakayama *et al.*² and Kent *et al.*³ claim that insulin itself is the primary autoantigen that initiates spontaneous type 1 diabetes in mice and humans, respectively, a result that could have implications for more effective prevention and therapy. However, I believe that this proposed immunological role of insulin may be undermined by the atypical responses of T cells to the human insulin fragment that are described by Kent *et al.*³.

Nakayama and colleagues² eliminated the normal insulin genes in a non-obese diabetic (NOD) mouse model and replaced them with an engineered gene encoding insulin that is hormonally active but fails to stimulate T-cell responses. The mice did not develop insulin antibodies, inflammation of β -islet cells (insulinitis), or diabetes. This important and surprising finding indicates that the absence of an immune response to a single fragment of the insulin B-chain, residues 9–23, could be sufficient to prevent onset of a disease that might be expected to be mediated by immune reactivity to many other potential β -cell autoantigens.

How could this happen? T cells recognize, and are activated by, antigenic peptide fragments bound to gene products of the major histocompatibility complex (MHC). $CD4^+$ T cells in the NOD mouse recognize the insulin B9–23 fragment complexed to the I-A^{B7} MHC class II molecule⁴; $CD8^+$ T cells recognize a related fragment, B15–23, in association with the H-K^d class I MHC molecule⁵. The mutated insulin transgene used by Nakayama *et al.* encodes a molecule with a tyrosine-to-alanine replacement at residue 16: presumably this causes the B9–23 fragment to bind with insufficient avidity to this I-A^{B7} molecule and/or failure of the B15–23 peptide to bind to H-K^d, making these insulin fragments unrecognizable to the T-cell-receptor repertoire. Prevention of diabetes in these NOD mice therefore involved replacing an immunogenic and hormonally active insulin molecule with a non-immunogenic one.

Kent and colleagues³ recovered T-cell lymphocytes from the pancreatic-draining lymph

nodes of diabetic patients, cloned them and attempted to identify the protein antigen fragments recognized by these T cells. Remarkably, a substantial proportion (about 50%) of $CD4^+$ T cells recognized the amino-terminal fragment (residues 1–15) of the insulin A-chain. More surprising, however, is that the proliferative responses by these clones to the insulin peptide were at best minimal: stimulation indices were three- to sixfold; one of three clones gave an 80-fold increase in interleukin-13 production, whereas the other two produced interleukin-13 in amounts that were only three to six times above background; and very large amounts of peptide (1 mM, or about 1.6 mg ml⁻¹) were required to elicit these responses.

These seem to be valid T-cell responses in that they can be blocked by monoclonal antibodies against MHC class II peptides, but they are notably atypical in the amount of antigen required for stimulation and in the level of the resulting response. For example, this does not compare with the functional avidity of most myelin-specific $CD4^+$ T cells from patients with multiple sclerosis, which have EC₅₀ values (where EC₅₀ is the effective concentration of antigen that causes half-maximal response) in the range 1–50 μ M (ref. 6), corresponding to an antigen concentration that is 200–1,000-fold lower; also, a typical encephalitogenic $CD4^+$ T-cell clone in experimental allergic

encephalomyelitis, an autoimmune murine model for multiple sclerosis, requires a concentration of a myelin basic protein peptide fragment in the 300 nM range (about 300 ng ml⁻¹) for activation⁷, which is 5,000-fold lower.

The results of Kent *et al.* may indicate either that the native ligand for these pancreatic lymph-node T-cell clones uses a sequence other than the first 15 residues of the insulin A-chain and is a different nominal antigen that may be unrelated to the autoimmune process, or that these anti-insulin clones have a very low avidity. But in the latter case, where in the body would the required concentration of insulin peptide be available to activate these cells?

Darcy B. Wilson

Torrey Pines Institute for Molecular Studies,
San Diego, California 92121-1122, USA
e-mail: dbwilson@tpims.org

1. Castano, L. & Eisenbarth, G. S. *Annu. Rev. Immunol.* **8**, 647–679 (1990).
2. Nakayama, M. *et al.* *Nature* **435**, 220–223 (2005).
3. Kent, S. C. *et al.* *Nature* **435**, 224–228 (2005).
4. Abiru, N. *et al.* *J. Autoimmun.* **14**, 231–237 (2000).
5. Wong, F. S., Moustakas, A. K., Wen, L., Papadopoulos, G. K. & Janeway, C. A. Jr *Proc. Natl Acad. Sci. USA* **99**, 5551–5556 (2002).
6. Muraro, P. A., Pette, M., Bielekova, B., McFarland, H. F. & Martin, R. J. *Immunol.* **164**, 5474–5481 (2000).
7. Kumar, V. *et al.* *Proc. Natl Acad. Sci. USA* **92**, 9510–9514 (1995).

doi:10.1038/nature04423

IMMUNOLOGY

Kent *et al.* reply

Replying to: D. B. Wilson *Nature* **438**, doi:10.1038/nature04423 (2005)

Wilson discusses the avidity of the lymph-node T-cell clones we described¹ that recognize the insulin A1–15 fragment. He questions whether large enough concentrations of insulin would be available *in situ* for antigenic stimulation of T cells², given the amount of peptide required to stimulate the clones and their degree of responsiveness.

To evaluate the response of these insulin-reactive T cells generated without antigen stimulation, relative to that of human autoreactive T-cell clones generated in response to myelin basic protein (MBP), we have compared the proliferative response of the Ob.1A12 T-cell clone (which recognizes the MBP p85–99 epitope³) with that of our insulin-peptide-reactive T-cell clones (unpublished). The subject was

DR2- and DR4-positive and the clone was generated using autologous antigen-presenting cells expressing both human leukocyte antigen (HLA) class II alleles. The response with DRB1*1501 as a restricting element was stronger, requiring an EC₅₀ of 0.1–10 μ M MBP p85–99 peptide, whereas the response with DRB1*0401 required an EC₅₀ of about 100 μ M peptide; this was similar to the EC₅₀ response of insulin-peptide-reactive T cells restricted by DRB1*0401 (ref. 1). (EC₅₀ is the effective concentration of antigen that causes half-maximal response.)

It is known that human T-cell clones in long-term culture (grown for years, for example) vary in their amount of proliferation and cytokine secretion (our unpublished observa-

tions), which is influenced by their activation state: therefore, the actual proliferative response may not necessarily reflect antigen sensitivity. Perhaps more important are data we have obtained since our original report: reduction of the three cysteine residues in the insulin A-chain 1–15 peptide (to abolish its secondary and tertiary structure) markedly increases the sensitivity of the insulin-reactive CD4⁺ T cells, giving a significant response with 10 μ M insulin peptide (our unpublished observations). These insulin-reactive T-cell clones therefore behave like other autoreactive T-cell clones when recognizing denatured insulin antigen.

However, it is still not known whether the T-cell clones recognize naturally processed insulin epitopes *in vivo*. Although responses to MBP and to glutamic acid decarboxylase 65 (GAD65) are consistently detectable in peripheral blood^{3,4}, we never observe the same degree of proliferation in response to insulin as to these other, more sequestered antigens. We agree that a microbe or some other self-antigen could have been the primary trigger

for production of these insulin-reactive T-cell clones and that we may be observing a cross-reactivity that is pathogenic^{5–7}.

We next consider Wilson's question about the site in the body where the required concentration of insulin peptide would be available to activate these cells². Insulin in an islet-cell granule is present at about 240 mg ml⁻¹ (ref. 8), which is much higher than the concentrations of insulin peptide we used¹. If an insulin granule were to be phagocytosed by a dendritic cell, the concentration in the early endosomal compartment would be extremely high, although bound peptide would dissociate rapidly. Sufficiently high concentrations are therefore readily obtainable for dendritic cells carrying antigen to the draining lymph nodes, although the avidity of the T-cell clones for insulin is not known. Investigation of the receptors from these clonally expanded T cells from the draining lymph nodes of diabetic subjects should help in understanding the potential cross-reactivity that may exist in the induction of human autoimmune diseases.

David A. Hafler*, **Sally C. Kent***, **Yahua Chen***,

Lisa Bregoli*, **Sue M. Clemmings†**,
Bernhard Hering†, **Norma Sue Kenyon‡**,
Camillo Ricordi‡

*Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

e-mail: dhafler@rics.bwh.harvard.edu

†Department of Surgery, University of Minnesota, Minneapolis, Minnesota 55455, USA

‡Diabetes Research Institute, Cell Transplant Center, University of Miami, Miami, Florida 33136, USA

1. Kent, S. C. *et al. Nature* **435**, 224–228 (2005).
2. Wilson, D. B. *Nature* **438**, doi:10.1038/nature04423 (2005).
3. Ota, K. *et al. Nature* **346**, 183–187 (1990).
4. Viglietta, V., Kent, S. C., Orban, T. & Hafler, D. A. *J. Clin. Invest.* **109**, 895–903 (2002).
5. Wucherpfennig, K. W. & Strominger, J. L. *Cell* **80**, 695–705 (1995).
6. Ausubel, L. J., Kwan, C. K., Sette, A., Kuchroo, V. K. & Hafler, D. A. *Proc. Natl Acad. Sci. USA* **93**, 15317–15322 (1996).
7. Hemmer, B. *et al. Nature Med.* **5**, 1375–1382 (1999).
8. Hutton, J. C., Penn, E. J. & Peshavaria, M. *Biochem. J.* **210**, 297–305 (1983).

doi:10.1038/nature04424

Reconstruction of genetic circuits

David Sprinzak¹ & Michael B. Elowitz¹

The complex genetic circuits found in cells are ordinarily studied by analysis of genetic and biochemical perturbations. The inherent modularity of biological components like genes and proteins enables a complementary approach: one can construct and analyse synthetic genetic circuits based on their natural counterparts. Such synthetic circuits can be used as simple *in vivo* models to explore the relation between the structure and function of a genetic circuit. Here we describe recent progress in this area of synthetic biology, highlighting newly developed genetic components and biological lessons learned from this approach.

By taking apart an old clock, you could probably come up with a pretty good guess at how it works. But a more concrete understanding of the clock mechanism might be obtained by designing and building one's own clock out of similar parts. Contemporary biology presents us with similar reverse-engineering problems. For example, *Drosophila* cells contain a circadian clock that oscillates with a 24-h rhythm and self-synchronizes to the day/night cycle. Using genetic and biochemical techniques, researchers have isolated genes and proteins involved in interlocked feedback loops of gene expression^{1,2} (Fig. 1b) that are necessary for clock function. However, many fundamental questions remain difficult to answer: what sets the period of the oscillation, how does the clock operate reliably in diverse cellular conditions, and what features of its design are responsible for its reliable operation? To gain insight into such questions one could design and build new clock circuits, using similar genes and proteins, and study their dynamics in the organism. In fact, several synthetic genetic clocks have now been constructed in bacteria^{3–5} (Fig. 1c). These circuits are much simpler than the *Drosophila* clock. They fail to operate as reliably, but they provide a proof of principle for a synthetic approach to understanding genetic circuits.

As with the clock, many compelling biological questions centre on how interactions among genes and proteins, forming genetic circuits, give rise to specific cellular functions. Genetic and biochemical techniques have successfully identified many circuit components and their interactions. However, in many cases knowledge about these components and their interactions is not sufficient to explain the circuit mechanism. What is missing from the circuit diagram? There are several possible deficiencies: one is that the diagram may be incomplete—interactions may have been missed. The opposite problem also exists: the diagram may be too complete, in the sense that it contains interactions that are not actively involved in the process being investigated (for example, if one of the proteins is not expressed under the relevant conditions). Another problem is ignorance of the effective rules by which proteins and genes interact. For example, *in vivo* values of kinetic parameters such as affinities, binding and degradation rates, and so on, are generally unknown. Finally, the intracellular environment is intrinsically 'noisy', and small copy numbers of molecular species limit the predictability of biochemical reactions⁶. Taken together, these problems reduce our confidence in the combined understanding we get from perturbations, measurements and mathematical modelling.

A reconstructive approach to genetic circuits may offer unique

insight into their underlying mechanisms. In this approach, one constructs synthetic replicas of natural genetic circuits out of well-characterized elements, such as genes, proteins, regulatory sequences, and so on, and observes their dynamics in living cells (Fig. 1a). This scheme offers several advantages. First, one can test the sufficiency of an arbitrary circuit for generating a particular function. Second, one may study the circuit mechanism without impairing cellular functions or inducing downstream consequences. Third, different circuit designs with similar functions can be directly compared to determine their relative advantages and disadvantages⁷. Such synthetic circuit reconstruction is made possible by the technologies of molecular biology such as construction of recombinant DNA molecules. Its future progress will benefit from further advances such as recent improvements in longer DNA synthesis⁸.

In this way, synthetic genetic circuits can function as physical models of natural genetic circuits. The use of physical models to understand natural phenomena is common throughout the sciences in general and biology in particular. Linus Pauling discovered the α -helix structural motif by building physical ball-and-stick models of proteins⁹. In molecular biology, the synthesis of an artificial chromosome in yeast proved the necessity and sufficiency of particular combinations of sequence elements for chromosome function¹⁰. Most similarly, in the field of protein design, proteins can be designed *de novo* based on principles and motifs discovered in natural proteins^{11,12}. For example, understanding of protein–protein interactions is facilitated by design of synthetic interacting proteins¹².

We describe here how synthetic biology can address biological questions at the level of genetic circuits. We review some tools developed for synthetic biology and a few specific experiments that use these tools to answer fundamental biological questions. Finally, we discuss future directions of this approach. Owing to space limits, we confine ourselves to a small subset of the much larger, and rapidly growing, field of synthetic biology, reviewed elsewhere^{13–18} and in this issue¹⁹.

Modular circuit components

In order to build synthetic circuits, one needs genetic components that are well characterized, modular (that is, function similarly in different systems) and act independently of other cellular processes. Early synthetic biology experiments focused on transcriptional regulation components because they are relatively well understood and easy to reconfigure. For example, repressors were used to create feedback loops of various sizes in *Escherichia coli* in

¹California Institute of Technology, Division of Biology and Department of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA.

order to understand noise suppression²⁰, bistability^{21,22} and oscillations³ in circuits of one, two and three repressors, respectively. Similarly, synthetic cascades without feedback provided information on delays²³, noise propagation^{6,24} and sensitivity²⁵. More recently, some of these transcriptional circuit designs have been created and analysed in mammalian systems using newly developed transcriptional regulators^{26,27}. In one case, a synthetic bistable switch was shown to operate in a mouse²⁶.

Nevertheless, it is clear that many natural circuits are fundamentally non-transcriptional. An amazing example is the cyanobacterial circadian clock, the operation of which depends on protein phosphorylation but can be independent of transcription and translation (in contrast to its counterpart in *Drosophila*)²⁸. Thus, it is critical to appropriate other interaction mechanisms, such as protein modification, regulated degradation, and so on, for use in synthetic circuits.

One class of such non-transcriptional regulatory control mechanism involves *trans*-acting RNA molecules, such as microRNAs. Regulation by RNA offers promising applications in synthetic biology because of the inherent versatility of its sequence-based targeting mechanism²⁹. Recently, Bayer and Smolke³⁰ reported the development of designed 'anti-switches' composed of a modular ligand-binding (aptamer) domain joined to a regulatory domain that inhibits translation of specific target mRNAs in a ligand-dependent fashion (Fig. 2a). Isaacs *et al.*³¹ described 'cis-repressed' riboregulatory molecules containing a translation-repressing hairpin RNA structure. Complementary *trans*-acting RNA molecules relieve inhibition by the hairpin, allowing gene expression. Both types of

components should find applications in synthetic circuits that require combinatorial control of gene expression.

Small molecules such as hormones and metabolites carry out many cellular functions—can they be used in synthetic genetic circuits? Recently, Fung *et al.*⁵ created the 'metabolator', a synthetic circuit in which transcriptional control of metabolic enzymes generate oscillations between two metabolite pools: acetyl CoA versus acetyl phosphate and acetate (Fig. 2b). The metabolic enzymes that interconvert these two metabolites are under the transcriptional control of acetyl phosphate-sensitive regulators. On a larger scale, metabolic enzymes from plant, yeast and *E. coli* were arranged in synthetic pathways to produce amorpha-4,11-diene³², a precursor of the highly effective anti-malarial drug artemisinin. Synthetic metabolic pathway construction has the potential to enable synthesis of medically important compounds and may also shed light on fundamental issues of pathway design.

Small diffusible molecules naturally used for intercellular communication have also been appropriated for synthetic circuits. For example, acyl-homoserine lactone (AHL), a signalling molecule in diverse bacteria³³, was recently³⁴ used to implement a simple multicellular patterning system in *E. coli*, loosely analogous to a morphogen gradient³⁵ (Fig. 2c). In this work, AHL was produced by a 'sender' strain localized to the centre of a Petri dish. A second set of 'receiver' strains contained synthetic transcriptional circuits that combine positive and negative responses to AHL and thereby serve as 'band-pass' detectors. As a result, the fluorescent output gene was expressed only in cells within a specific range of distances from the source (that is, in a ring). Owing to an unanticipated design feature of the circuit,

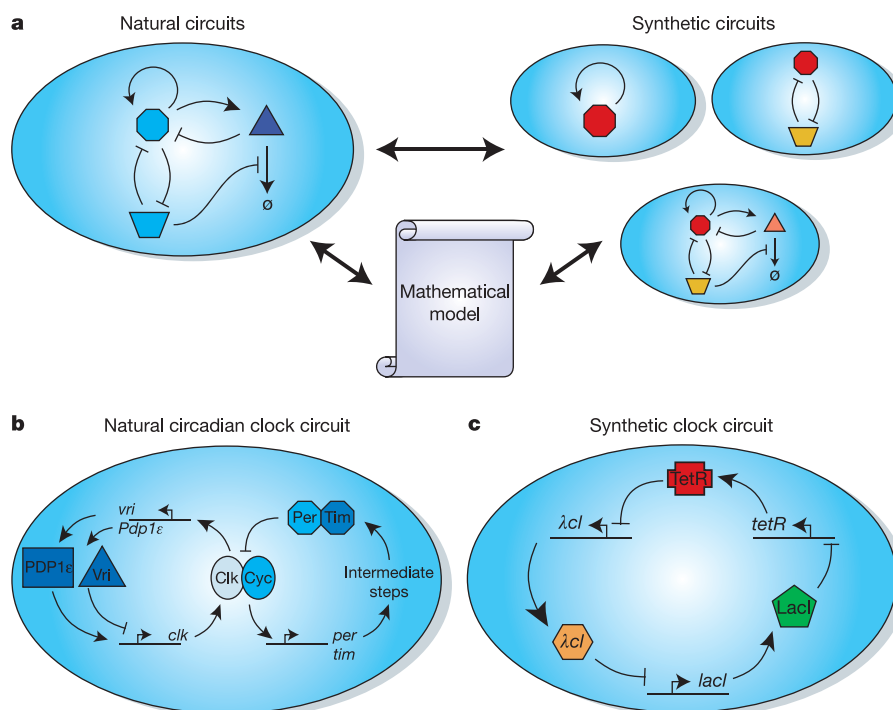


Figure 1 | Natural and synthetic genetic circuits. **a**, The synthetic biology paradigm. Genetic circuits are composed of interacting genes and proteins (blue shapes, top left). The pointed and blunt arrows represent positive and negative regulation, respectively. Synthetic circuits (top right) based on the natural circuit can be constructed from well-characterized components (red and orange shapes) with similar regulatory effects to form similar or simplified circuits. The dynamics of these synthetic replicas can be compared to the natural system as well as to mathematical models. Analysis of natural circuits, synthetic replicas and models together can help us understand mechanisms used by natural systems. **b**, The *Drosophila* circadian clock is an actively studied natural clock circuit (simplified scheme based on Cyran *et al.*²). It contains a negative feedback loop in which Per and

Tim, after a delay, repress their own production (via Clk/Cyc, right loop). Interlocked with this negative feedback loop is another loop involving Vri and PDP1ε (left loop). The diagram is highly simplified and many details of the process have been omitted, including post-translational modifications, nuclear transport and active degradation. **c**, The 'repressilator' is a simple synthetic clock circuit consisting of a three-component negative feedback loop that operates in *E. coli*². The three-element loop provides a delayed negative feedback on all components and permits oscillations. In this sense, it models the generation of oscillations by delayed negative feedback. However, as can be seen from the figure this simple synthetic circuit differs markedly from the natural circadian clock in both complexity and design.

the position of this ring did not vary over time as the signal diffused outward from its source. This effect was attributed to a dynamic delay within the circuit: one transcription factor had to be diluted by cell growth to allow expression of the reporter. Thus, the authors uncovered a design strategy for creating stationary patterns from a diffusive gradient that grows with time. We are not aware of evidence for this mechanism in natural morphogen systems; however, similar problems may appear in development, and it will be exciting to see whether synthetic circuits can address them.

One concern when building synthetic genetic circuits is the effects that they might have on endogenous cellular functions. Ideally, one would like circuits to operate independently of the rest of the cell, but the limits to such independence remain unclear. Recently, a step in this direction was taken by Rackham and Chin³⁶, who generated a set of 'orthogonal' ribosome-mRNA pairs. They used a novel selection-counterselection scheme to select three specifically interacting pairs of ribosome binding sites and their complementary regions on the 16S rRNA (Fig. 2d). These orthogonal ribosomes represent partially independent protein production lines. Their approach might be extended to other steps in gene expression, for example, by creating orthogonal sigma factors for prokaryotic transcription. Circuits based on such orthogonal systems may provide better understanding of the natural regulation of global processes like transcription and translation.

Lessons learned from synthetic circuits

The tools above, and others, will facilitate the development of novel synthetic circuits, but the ultimate value of this approach will depend on what general principles it reveals about circuit design. So far, several important lessons have emerged from the earliest results.

The intracellular environment differs in many ways from the idealized test tubes in which we imagine biochemical reactions occurring and that form the basis for most mathematical models. For example, the cellular environment is highly variable. This variability arises in two ways. First, owing to the low copy number of many cellular molecules, stochastic effects in biochemical reactions can be significant. This stochasticity, or 'intrinsic noise', causes biochemical reactions to be partly unpredictable, and therefore more difficult to model. Second, individual cells differ from one another in many ways. This 'extrinsic noise' causes cell-cell differences in circuit dynamics³⁷. There is an interplay between the two types of noise: intrinsic noise in one cellular component, such as a transcription factor, may contribute to extrinsic noise in other components, such as its target genes. Synthetic biology experiments have been instrumental in probing both types of noise and revealing the strong effect they have on the function of genetic circuits.

Rosenfeld *et al.*³⁸ used time-lapse movies to analyse a synthetic cascade in which a fluorescent repressor protein and its target gene were monitored simultaneously in individual cells (Fig. 3a). Their

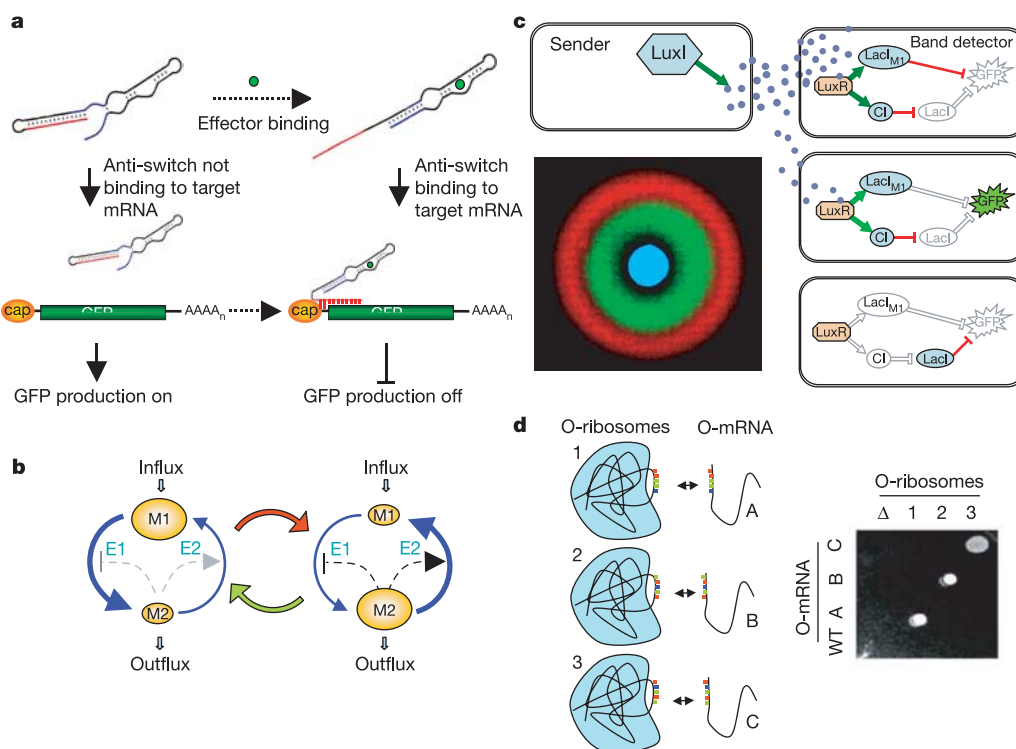


Figure 2 | Modular components in synthetic circuits. a, RNA 'anti-switches'. These molecules contain an aptamer domain that binds to a specific effector (small molecule ligand), as well as a sequence recognition domain complementary to a target mRNA. Binding of the effector changes the conformation of the switch and enables control of target gene expression³⁰. **b**, The 'metabolator'. This synthetic oscillatory circuit relies on regulated metabolic fluxes. Oscillations occur between two pools of metabolites, here labelled M1 and M2 (see text). M2 indirectly regulates the metabolic enzymes E1 and E2, corresponding to phosphate acetyltransferase and acetate kinase. Two extreme states of the oscillator are shown in the diagram. On the left, the system has lots of M1 and little M2. In this state, the enzyme activities are small, causing substantial net flux from M1 to M2. On the right is the opposite state in which M2 levels are high and M1 levels are low. Now, E1 and E2 levels increase, reversing the net flux⁵. **c**, Formation of

spatial patterns by a synthetic multicellular circuit. Mixed populations of engineered *E. coli* strains are placed on a Petri dish. Sender cells, localized to the centre (top left), emit AHL, which diffuses away. Band detector strains throughout the plate (right) respond to specific ranges of AHL using a synthetic band detector circuit. The result is concentric ring patterns of gene expression (bottom left), in which two different detector strains (green and red) turn on at specific distances from the source³⁴. **d**, Synthesis of orthogonal ribosome-mRNA pairs. The 16S ribosomal subunit and a target ribosome-binding site sequence were coevolved to recognize one another specifically and uniquely (left). A chloramphenicol resistance gene was subcloned together with each of the evolved ribosome-binding sites and co-transformed with each evolved, and one wild-type, 16S subunit, to produce a 'matrix' of strains. As shown on the right, only strains containing matched pairs are able to grow in media containing high levels of chloramphenicol³⁶.

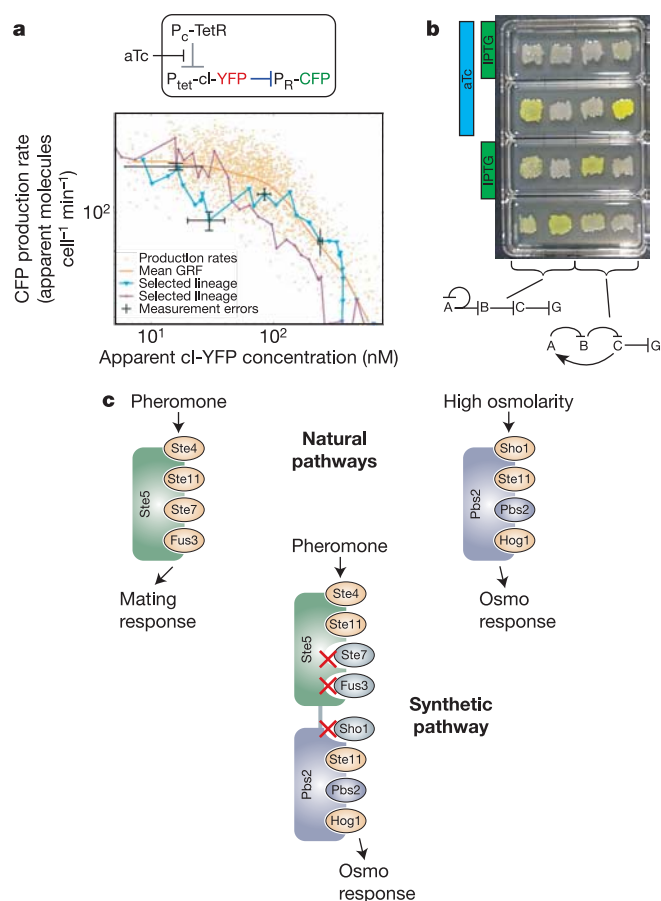


Figure 3 | Lessons from synthetic biology. **a**, Noise in gene regulation⁶. A synthetic transcriptional cascade was constructed to analyse the gene regulation function (GRF) in single cells (top panel). The GRF is the relation between the protein production rate (y axis in graph) and the concentration of its repressor protein (x axis). The transcriptional cascade comprises reporter gene, cyan fluorescent protein (CFP), under a λ P_R promoter, which can be repressed by CI repressor fused to yellow fluorescent protein (YFP). The CI-YFP fusion protein is under the control of TetR repressor (repression of TetR can be removed through addition of aTc (anhydrotetracycline); TetR is under a constitutive promoter P_C). Each orange point in the graph (bottom panel) represents a measurement made in a single cell at a given time. The average GRF is given by the orange line. Measurements from two selected lineages are indicated in cyan and magenta. The data show significant fluctuations in the GRF that vary slowly, with a timescale of approximately one cell cycle. These results show that the GRF is not well defined at the single cell level. **b**, Exploring the functional potential of transcriptional circuits⁴⁰. Three transcriptional regulators and five promoters were rearranged to produce a small library of random genetic circuits, four of which are shown here (columns). For each circuit, the expression of a green fluorescent reporter gene (G) was measured with and without each of two extracellular inducers (IPTG and aTc). Circuits with four different output functions (yellow–green fluorescence versus four indicated input conditions) are shown here patched onto agar plates; however, each pair shares a single circuit structure, shown below. In these diagrams, A, B and C refer to transcriptional regulatory proteins. Note that different functions are seen with the same diagram (columns 1–2 and 3–4), and similar (but not identical) functions are produced by different diagrams (columns 1 and 4)⁴⁹. **c**, Rewiring signalling pathways. The pheromone and osmolarity response pathways in yeast (top) use protein scaffolds (Ste5 and Pbs2) to avoid crosstalk through the shared component Ste11. By constructing a modified fusion protein of Pbs2 and Ste5, the authors constructed a rewired pathway in which cells produce osmolarity stress responses after pheromone induction⁴².

results showed how noise and gene regulation combine to determine the instantaneous rate of expression of a typical gene. The experiments assessed the timescale and amplitude of noise, in the regulation of a specific gene, and showed that the extrinsic component of the noise makes a dominant contribution to the total noise. They also revealed that most fluctuations do not ‘average out’ over a cell cycle, and therefore the response curve of a repressor–promoter system is not well defined at the single cell level. In related work, Pedraza and van Oudenaarden²⁴ constructed a linear cascade of three repressors, together with three corresponding fluorescent protein reporters. They showed how fluctuations propagate in a transcriptional cascade. This work used a previously developed theoretical framework³⁹ to provide an integrated analysis of noise at the circuit level. It remains to be seen how cells and human genetic circuit designers might use such fundamentally noisy components to make reliable circuits.

Synthetic biology relies on the rearrangement of modular elements to create novel circuits. Rather than designing a specific circuit, it is also possible to create combinatorial libraries of many possible circuits. What can such libraries teach us about the computational power and functional diversity of simple biological systems?

Recently, Guet *et al.*⁴⁰ addressed this question by creating a synthetic circuit ‘slot machine’ in which each of three transcriptional regulatory genes was controlled by one of five possible promoters (Fig. 3b). The resulting circuit library was screened for its response to two different chemical inputs that interacted with two of the repressors. A surprisingly large fraction of the diverse circuit designs generated computationally interesting functions, such as NAND and NOR. In this example, the map of known interactions among the components did not uniquely determine the circuit function. Rather, examples were found in which the same circuit architecture produced different functions, as well as others in which the same function was encoded by circuits with different architectures. This synthetic biology experiment demonstrates what a rich functional diversity can be obtained with even a relatively small number of components.

Similarly, Dueber *et al.*⁴¹ recently explored analogous structure–function relations at the level of protein domains. They built a library consisting of a kinase domain fused to various ligand-binding regulatory domains placed in different positions within the protein. They determined the activity of the kinase in the presence and absence of two ligands specific for domains contained in each protein. They found that a large diversity of different functions could be encoded in individual proteins by recombining modular protein domains. The implications at the protein level are similar to those at the circuit level: recombination of modular units can rapidly explore the space of possible functions. It remains to be seen what limits the functional spectrum in these cases.

A final lesson from synthetic circuits involves fidelity in signal transduction. Park *et al.*⁴² used a synthetic approach to study how signal fidelity is maintained within pathways that share components. In yeast, both the mating pheromone and osmolarity stress response pathways signal through the mitogen-activated protein (MAP) kinase Ste11, yet they exhibit no crosstalk. This insulation is accomplished by the protein scaffolds Ste5 and Pbs2 that bind Ste11 together with other pathway-specific components (Fig. 3c). Even though Ste11 participates in two pathways (interacts with two scaffolds), it transduces a signal while in a complex with specific upstream and downstream components. Park *et al.*⁴² replaced natural tethering interactions with synthetic alternatives that were sufficient to tether together the different components of the MAP kinase cascade and allow signal propagation. To show the sufficiency of tethering, they created a synthetic hybrid scaffold consisting of a modified fusion protein of Pbs2 and Ste5. This construct effectively re-wired the signalling pathway so that pheromone induced osmolarity stress responses. Here, the synthetic approach involves rewiring, rather than reconstruction per se, and demonstrates that the

principal requirement for signalling through a pathway is proximity of components. It also suggests a simple mechanism by which signal transduction pathways might evolve new capabilities.

Challenges for synthetic circuit design

Constructing a functional synthetic circuit requires assembling diverse genetic elements and getting them to work together. In general, combining disparate components requires the tuning of biochemical parameters such as affinities or rate constants, which is often difficult to implement in biological circuits^{3,25}. Characterization of a component may be valid in one context (locus, plasmid, strain, environment, and so on), but not in others. How can one design an operating circuit given these limitations? Several strategies have been applied. First, the use of tunable elements, such as transcription factors derived from *tetR*^{43,44}, allows external control over some parameters. Second, one can screen libraries of mutated components, or apply directed evolution in the laboratory, to optimize parameters⁴⁵. A third strategy is to use robust circuit designs that are inherently insensitive to unknown or variable parameters. Such designs are particularly interesting because they may have been selected by evolution for the very same property⁴⁶.

A related challenge is computational modelling of genetic circuits. Modelling is essential both for analysis of natural systems and also for design of synthetic ones. However, several problems complicate its application to cellular circuits. These include parameter sensitivity, the lack of effective rules to simplify complex circuits, and the difficulty of incorporating extrinsic noise. Because synthetic circuits are simpler and better characterized than their natural counterparts, they will probably offer ideal test systems to develop and refine models. The results should apply both to natural and synthetic circuits.

Future directions

What are the goals of the synthetic circuit paradigm outlined here? One is to better understand natural circuits by building minimal replicas of those circuits, observing their dynamics *in vivo*, and comparing them to one another and to their natural counterparts. The synthetic circuits presented above are highly simplified. However, as we gain confidence and expertise in our ability to build, model and analyse these circuits, we will be able to construct replicas of greater verisimilitude. Possible natural circuits that could be investigated this way include decision making in response to stress and DNA damage, as in the natural p53/mdm2 circuit⁴⁷, differentiation in response to extracellular signals, as in oocyte maturation⁴⁸, and regulated temporal oscillations, as in the cell cycle and circadian clock^{1,10}. Circuits that use the intrinsically noisy nature of the cell to create probabilistic behaviours are particularly compelling examples.

A second goal is to discover what other, non-natural, circuit designs are possible given realistic biological components, and which of those operate reliably *in vivo*. This will be achieved by building and characterizing a variety of alternative circuit designs in living cells. In this way, one may ask what advantages naturally evolved circuits have over synthetic ones. For example, the synthetic clock designs described earlier have not been discovered to occur in nature, suggesting that natural designs may confer better performance. At the same time, non-natural designs may prove useful for biotechnology applications³².

Perhaps the most intriguing problem is how a circuit operates in the context of a complete organism. There are no dotted lines inside the cell isolating circuits from one another. The ultimate test for this synthetic approach is to delete natural circuits and replace them with synthetic counterparts within organisms. This will require synthetic circuits to interface with the rest of the cell. For example, by replacing the *Drosophila* circadian clock with synthetic versions we could learn more about the interaction of the circadian module with other functional subsystems in the organism. Even the most optimistic synthetic biologist would expect such circuits to be less functional

than their natural counterparts. However, perhaps at this stage one can learn more by putting together a simple, if inaccurate, pendulum clock than one can by disassembling the finest Swiss timepiece.

- Hardin, P. E. The circadian timekeeping system of *Drosophila*. *Curr. Biol.* **15**, R714–R722 (2005).
- Cyran, S. A. et al. *vrrile*, *Pdp1*, and *dClock* form a second feedback loop in the *Drosophila* circadian clock. *Cell* **112**, 329–341 (2003).
- Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
- Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behaviour in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
- Fung, E. et al. A synthetic gene–metabolic oscillator. *Nature* **435**, 118–122 (2005).
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
- Alves, R. & Savageau, M. A. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics* **16**, 786–798 (2000).
- Tian, J. et al. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
- Judson, H. *The Eighth Day of Creation* (Simon and Schuster, New York, 1979).
- Ingolia, N. T. & Murray, A. W. The ups and downs of modeling the cell cycle. *Curr. Biol.* **14**, R771–R777 (2004).
- Bolon, D. N., Voigt, C. A. & Mayo, S. L. De novo design of biocatalysts. *Curr. Opin. Chem. Biol.* **6**, 125–129 (2002).
- Kortemme, T. & Baker, D. Computational design of protein–protein interactions. *Curr. Opin. Chem. Biol.* **8**, 91–97 (2004).
- Brent, R. A partnership between biology and engineering. *Nature Biotechnol.* **22**, 1211–1214 (2004).
- McDaniel, R. & Weiss, R. Advances in synthetic biology: on the path from prototypes to applications. *Curr. Opin. Biotechnol.* **16**, 476–483 (2005).
- Benner, S. A. & Sismour, A. M. Synthetic biology. *Nature Rev. Genet.* **6**, 533–543 (2005).
- Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
- Arkin, A. P. Synthetic cell biology. *Curr. Opin. Biotechnol.* **12**, 638–644 (2001).
- Pawson, T. & Linding, R. Synthetic modular systems—reverse engineering of signal transduction. *FEBS Lett.* **579**, 1808–1814 (2005).
- Endy, D. Foundations for engineering biology. *Nature* doi:10.1038/nature04342 (this issue).
- Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
- Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
- Kobayashi, H. et al. Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl Acad. Sci. USA* **101**, 8414–8419 (2004).
- Rosenfeld, N. & Alon, U. Response delays and the structure of transcription networks. *J. Mol. Biol.* **329**, 645–654 (2003).
- Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
- Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA* **102**, 3581–3586 (2005).
- Kramer, B. P. et al. An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnol.* **22**, 867–870 (2004).
- Kramer, B. P. & Fussenegger, M. Hysteresis in a synthetic mammalian gene network. *Proc. Natl Acad. Sci. USA* **102**, 9517–9522 (2005).
- Nakajima, M. et al. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. *Science* **308**, 414–415 (2005).
- Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
- Bayer, T. S. & Smolke, C. D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnol.* **23**, 337–343 (2005).
- Isaacs, F. J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnol.* **22**, 841–847 (2004).
- Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnol.* **21**, 796–802 (2003).
- Henke, J. M. & Bassler, B. L. Bacterial social engagements. *Trends Cell Biol.* **14**, 648–656 (2004).
- Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
- Ephrussi, A. & St Johnston, D. Seeing is believing: the bicoid morphogen gradient matures. *Cell* **116**, 143–152 (2004).
- Rackham, O. & Chin, J. W. A network of orthogonal ribosome–mRNA pairs. *Nature Chem. Biol.* **1**, 159–166 (2005).
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).

38. Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323**, 785–793 (2002).
39. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
40. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1466–1470 (2002).
41. Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. Reprogramming control of an allosteric signalling switch through modular recombination. *Science* **301**, 1904–1908 (2003).
42. Park, S. H., Zarrinpar, A. & Lim, W. A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061–1064 (2003).
43. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1–I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
44. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl Acad. Sci. USA* **89**, 5547–5551 (1992).
45. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99**, 16587–16591 (2002).
46. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
47. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
48. Xiong, W. & Ferrell, J. E. Jr A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426**, 460–465 (2003).
49. Guet, C. C. *Tinkering with Bio-molecular Networks* Thesis, Princeton Univ. (2004).

Acknowledgements Synthetic biology work in the laboratory is supported by a Burroughs-Wellcome CASI award, The HFSP program and the Searle Scholars program. D.S. would like to acknowledge the support of the Yad Hanadiv Foundation and the CBCD at Caltech. We would like to thank J. W. Chin, C. Guet, S. Leibler, J. Liao, W. Lim, C. Smolke and R. Weiss for contributing figures. We are grateful to R. Kishony, E. Sprinzak, C. Dalal, G. Suel and members of the laboratory for discussions and comments.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to M.B.E. (melowitz@caltech.edu).

REVIEWS

Foundations for engineering biology

Drew Endy¹

Engineered biological systems have been used to manipulate information, construct materials, process chemicals, produce energy, provide food, and help maintain or enhance human health and our environment. Unfortunately, our ability to quickly and reliably engineer biological systems that behave as expected remains quite limited. Foundational technologies that make routine the engineering of biology are needed. Vibrant, open research communities and strategic leadership are necessary to ensure that the development and application of biological technologies remains overwhelmingly constructive.

Please complete one of the following projects in the next hour: write down the DNA sequence that programmes a biofilm to take a photograph and perform distributed edge-detection on the light-encoded image; or, the DNA sequence that encodes a ring oscillator that works inside yeast; or, the DNA sequence that programmes any mammalian cell to count up to 256 in response to a generic input signal; or, the DNA sequence that programmes any prokaryote to produce 25 g l⁻¹ artemisinic acid. Alternatively, describe in convincing detail the starting materials, experimental screens and genetic selections that could be used to evolve biological systems to perform these tasks... Time's up!

The abovementioned applications of synthetic biology could find uses in the construction of templated surfaces for nanoscale materials fabrication, the characterization of physical performance limits for eukaryotic gene expression, the study and control of cell division, animal development, ageing and cancer by modest amounts of genetically encoded memory and logic, or the cost-effective production of an anti-malaria drug precursor, respectively. Furthermore, each application is physically plausible, or is the direct extension of an already demonstrated result¹⁻³. Yet, each project would be a unique, expert-driven research problem, with uncertain times to completion, costs and probabilities of success. For example, construction of the first genetically encoded 'ring oscillator' in bacteria took two of the world's best biophysicists at least one year of research^{2,4}. Constructing a second genetically encoded ring oscillator would probably cost another group as much effort. In contrast, any competent electrical engineering student could build many working electronic ring oscillators in under one hour.

In 1978, Szybalski and Skalka wrote⁵, "The work on restriction nucleases not only permits us easily to construct recombinant DNA molecules and to analyse individual genes, but also has led us into the new era of 'synthetic biology' where not only existing genes are described and analyzed but also new gene arrangements can be constructed and evaluated." Twenty-seven years later, despite tremendous individual successes in genetic engineering and biotechnology⁶⁻⁸, why is the engineering of useful synthetic biological systems still an expensive, unreliable and *ad hoc* research process?

The first possibility is that we don't yet know enough about biological systems, or that biological systems are too complex to reliably engineer, or both. For example, some descriptions of natural biological systems are notoriously complex^{9,10}. The large number of unique functional components combined with unexpected interactions among components (for example, pleiotropy) makes it hard to imagine that we might reliably engineer the behaviour of complex biological systems. Furthermore, it is possible that the designs of

natural biological systems are not optimized by evolution for the purposes of human understanding and engineering¹¹. Thankfully, these concerns are best evaluated by attempting to surmount them¹².

The second possibility is that the engineering of biology remains a research problem because we have never invented and implemented foundational technologies that would make it an engineering problem. Stated plainly, the engineering of biology remains complex because we have never made it simple (T. F. Knight, personal communication). As above, the practicality of making biological engineering simple can best be evaluated by attempting to make it simple. Success would help to "create the discipline of synthetic biology: an engineering technology based on living systems"¹³. Failures would directly illuminate and help prioritize the most relevant gaps in our current understanding of natural living systems, and suggest how we might best eventually come to understand and apply nature's original technology¹⁴.

What and why is synthetic biology?

The recent and ongoing interest in 'synthetic biology' is being driven by at least four different groups: biologists, chemists, 're-writers' and engineers. Briefly, for biologists, the ability to design and construct synthetic biological systems provides a direct and compelling method for testing our current understanding of natural biological systems^{4,15}; disagreements between expected and observed system behaviour can serve to highlight the science that is worth doing. For chemists, biology is chemistry, and thus synthetic biology is an extension of synthetic chemistry; the ability to create novel molecules and molecular systems allows the development of useful diagnostic assays and drugs, expansion of genetically encoded functions, study of the origins of life, and so on¹⁶. For 're-writers', the designs of natural biological systems may not be optimized for human intentions (for example, scientific understanding, health and medicine); synthetic biology provides an opportunity to test the hypothesis that the genomes encoding natural biological systems can be 're-written', producing engineered surrogates that might usefully supplant some natural biological systems¹¹. Finally, for engineers, biology is a technology; building upon past work in genetic engineering, synthetic biology seeks to combine a broad expansion of biotechnology applications with—as the focus of this article—an emphasis on the development of foundational technologies that make the design and construction of engineered biological systems easier.

Foundations for engineering biology

Many times over, individuals and groups have adapted and applied different resources from nature to the service of human needs such as

¹Division of Biological Engineering, Massachusetts Institute of Technology, Room 68-580, Koch Biology Building, 31 Ames Street, Cambridge, Massachusetts 02139, USA.

shelter, food, health and happiness; notably, natural resources are limited while our needs, in aggregate, may be unbounded. In this context, we should attempt to develop foundational technologies that make it easier and more efficient to satisfy human needs. For example, in the developed world, the design and construction of buildings is now a routine and reliable process. The success of the building process depends on (1) the existence of a limited set of predefined, refined materials that can be delivered on demand and that behave as expected, (2) generally useful rules (that is, simple models) that describe how materials can be used in combination (or not), and (3) skilled individuals with a working knowledge and means to apply these rules. Biology itself is a natural resource that can be further adapted to help satisfy human needs. What foundational engineering technologies would best enable the routine design and construction of useful synthetic biological systems?

Today, four challenges that greatly limit the engineering of biology are (1) an inability to avoid or manage biological complexity, (2) the tedious and unreliable construction and characterization of synthetic biological systems, (3) the apparent spontaneous physical variation of biological system behaviour, and (4) evolution. In considering how best to address these engineering challenges, one practical starting point is to consider past lessons from when other engineering disciplines emerged from the natural sciences. Are any past lessons relevant to the engineering of biology today? For example, could we usefully consider adapting or extending ideas from structural engineering to synthetic biology?

The three past ideas that now seem most relevant to the engineering of biology are standardization, decoupling and abstraction (see the Acknowledgements for a list of people who have helped to contribute to these ideas). To be clear, these ideas and my prioritization of them could be wrong; I would explicitly encourage the widespread invention and discussion of alternative ideas. However, as introduced below, the successful development of foundational technologies based on these ideas would greatly ameliorate the first three engineering challenges posed above. The fourth challenge, evolution, is largely unaddressed within past engineering experience (discussed below).

Standardization

Standards underlie most aspects of the modern world¹⁷. Railroad gauges, screw threads, internet addresses, 'rebar' for reinforcing concrete, gasoline formulations, units of measure, and so on. In the science of biology, a number of useful standards have already arisen around the 'central dogma' that defines the core operations of most natural biological systems, and in response to the development

of widely practiced methods that generate significant amounts of data. As representative examples, standards of varying utility now exist for DNA sequence data and genetic features¹⁸, microarray data¹⁹, protein crystallographic data²⁰, enzyme nomenclature²¹, systems biology models²² and restriction endonuclease activities²³. However, the biological engineering community has yet to develop formal, widely used standards for most classes of basic biological functions (for example, promoter activity), experimental measurements (for example, protein concentrations) and system operation (for example, genetic background, media, growth rate, environmental conditions, and so on).

Tremendous costs accrue owing to the lack of standards in biological engineering. For example, one would-be biological engineer's 'strong' ribosome binding site may be merely middling to another. As a second example, one group's genetic toggle switch may work in Jacques Monod's *Escherichia coli* strain JM2.300 (which may be the same as Sydney Brenner's *E. coli* strain XAO) when the cells are grown in Luria broth²⁴, while another group's synthetic genetic oscillatory network will work in Malcolm Casadaban's *E. coli* strain MC4100 when the cells are grown in supplemented minimal media². Could another group directly combine these two systems to produce a working 'toggle-lator'²⁵? The obvious engineering cost is that we don't know—it's a research project; but also note that it costs approximately 30 minutes just to track down all of this background information.

Looking forward, the biological engineering community would benefit from the development of technologies and the promulgation of standards that support the definition, description and characterization of the basic biological parts²⁶, as well as standard conditions that support the use of parts in combination and overall system operation. One very preliminary example of such work is emerging as the Registry of Standard Biological Parts (Fig. 1). However, beyond the technology itself, standards are also needed to support a vibrant, constructive and responsible community of biological engineers. For example, legal standards are needed to define means by which large collections of parts encoding basic biological functions, from a myriad of sources, can be easily shared and used in combination to realize many applications. Current legal standards such as uniform Materials Transfer Agreements (MTAs), which define the terms that allow researchers at different organizations to share biological materials, can take several weeks to execute, and typically do not resolve application rights. Meanwhile, direct chemical synthesis of up to ~10,000-base-pair DNA fragments is a widely available commercial service (as entering "gene synthesis" into an internet search engine will confirm); for cases in which the DNA sequence

Parts Catalog Click on the icons below to see parts by category. [more...](#)

Regulatory Reporter Inverter RNA Protein Generator Tag Primer Other Plasmid Cell Strain

RBS CDS Terminator Composite Cell-Cell Signalling Measurement

Web Site Update

The Registry web site has been moved from rosalind.csail.mit.edu to parts2.mit.edu. In addition, a few functions have been added:

- A BioBrick version of Blast compares sequences to parts in the Registry.
- DNA repositories keep track of the location of parts in Registry or (eventually) local freezers.

The new part viewer and editor is now available. It presents more data and allows in-place editing. The "User Experience" section of the part viewer now has some wiki features. Members of any moderated Registry group may edit the contents of that field. In the future, this wiki capability will be extended to more fields in the Registry.

Educational Programs

The Registry supports design classes where students make simple systems from standard, interchangeable biological parts and operate them in living cells.

Thirteen schools are participating in the 2005 Intercollegiate Genetically Engineered Machine competition (iGEM 2005). The schools are: Berkeley, Caltech, Cambridge, Davidson, ETH Zurich, Harvard, MIT, Oklahoma, Penn State, Princeton, Toronto, UCSF, and UT Austin.

Employment

The Registry is looking for full-time Technical Assistants. Please contact Staffing Services at MIT for details: [Technical Assistant](#).

Figure 1 | The Registry of Standard Biological Parts. This registry (<http://parts.mit.edu/>; modified image from the homepage shown), hosted by the MIT, provides free access to an open commons of basic biological

functions that can be used to programme synthetic biological systems²⁶. Anybody may contribute, draw upon, or improve the parts maintained within the Registry.

information encoding the desired part is freely available, it is often cheaper and faster to directly synthesize the part from scratch instead of executing an MTA. Thus, new low- or zero-cost legal strategies that facilitate and protect the exchange of the genetic information defining parts, and the use of parts in combination, are needed. As a second example, it seems prudent to develop community-wide standards for barcoding, signing and watermarking DNA in order to support the detection, identification and authentication of engineered biological systems, respectively²⁷.

Decoupling

Decoupling is the idea that it is useful to separate a complicated problem into many simpler problems that can be worked on independently, such that the resulting work can eventually be combined to produce a functioning whole. Two representative examples of decoupling include (1) building projects, which are often separated into architecture, engineering, construction, project management and inspection tasks, and (2) very-large scale integrated (VLSI) electronics, which is an engineering technology that only became practical once rules were worked out to enable the separation of chip design from chip fabrication²⁸.

Today, in biological engineering, many useful types of decoupling could be developed within the context of standardization (discussed above) and abstraction (discussed below). For example, one engineer might develop standard 'power supply and chassis' cells that provide known rates of nucleotides, amino acids and other resources to any engineered biological system placed within the cell, independent of the details of the system²⁹. Another engineer might develop technologies that enable the decomposition of complicated 'systems' into sets of independent 'devices' (discussed below). However, the simplest and most immediate type of decoupling is likely to be design

from fabrication. The decoupling of design and fabrication in biological engineering is being driven by recent and ongoing improvements in the process of DNA synthesis. Bulk DNA synthesis capacity appears to have doubled every 18 months or so for the last ten years; the commercial price of synthesis of long fragments of DNA (>500 base pairs) has continued to decrease by a factor of approximately two in each of the past three years³⁰—a trend that seems likely to continue over the next three years. More recently, downstream technology has been developed that allows for the automatic assembly of oligonucleotides and short DNA fragments into much longer molecules^{31–33}. Given gene and genome synthesis technology, some individuals can and should focus on designing useful pieces of DNA, while other individuals focus on building DNA; each group need only be expert in their respective tasks.

Abstraction

Natural biological systems seem wonderfully complex. New molecular mechanisms that regulate cellular behaviour are still being discovered³⁴ and 'exceptions to rules' abound³⁵. Given this context, how might we make routine the engineering of many-component biological systems that behave as expected?

One powerful technology for managing complexity is abstraction. Two forms of abstraction now seem worth exploring in biological engineering. First, the information describing biological functions might be organized across levels of complexity using abstraction hierarchies (Fig. 2). To be useful, biological engineering abstraction hierarchies must (1) allow individuals to work at any one level of complexity without regard for the details that define other levels, yet (2) allow for the principled exchange of limited information across levels (see Fig. 2 legend for more information). Second, the parts and devices that comprise engineered biological systems should probably

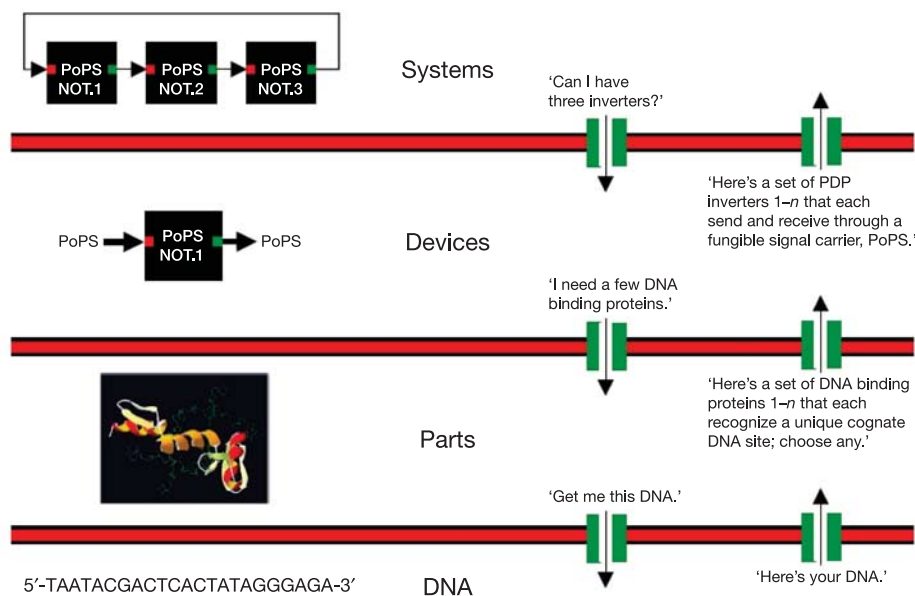


Figure 2 | An abstraction hierarchy that supports the engineering of integrated genetic systems. The purpose of an abstraction hierarchy is to hide information and manage complexity. Abstraction levels are listed ('DNA', 'Parts', 'Devices', 'Systems'). Here, 'DNA' is genetic material, 'Parts' are basic biological functions (for example, a DNA-binding protein), 'Devices' are any combination of 'Parts' that perform a human-defined function, and 'Systems' are any combination of 'Devices'. Abstraction barriers (red) block all exchange of information between levels. Interfaces (green) enable the limited and principled exchange of information between levels. Naïve exchanges are given in quotes. To be useful, individuals must be able to work independently at each level of the hierarchy. In synthetic biology, for example, parts-level researchers might need to know what sorts of parts that device-level researchers would like to use, how different types of

parts actually work (for example, atomic interactions between amino acids and the major groove of DNA), and how to order a piece of DNA. But, parts-level researchers should not need to know anything about phosphoramidite chemistry, how short oligonucleotides are assembled into longer, contiguous DNA fragments, or how a genetic oscillator works, and so on. As a second example, device-level researchers must define devices that use common signal carriers, and thus can be reused in combination (A. Che *et al.*, manuscript in preparation). Implicit in this hierarchy are formidable molecular engineering challenges; for example, engineering a set of 1,000 synthetic transcription factors based on zinc-finger-leucine-zipper chimaeras⁴⁵, each recognizing a unique cognate DNA binding site with >99% specificity (C. O. Pabo, personal communication).

be redesigned and built anew so that they are simpler to model and easier to use in combination. For example, it is often hard to predict the expressed protein level that will be produced by the novel combination of a natural transcription promoter, ribosome binding site (RBS) and open reading frame (ORF); combination-specific messenger RNA secondary structures created across RBS–ORF junctions might occlude ribosome loading and prevent an otherwise ‘strong’ RBS from initiating protein synthesis. Thus, in this example, it is worth exploring whether libraries of parts might be refined to produce common part–part junctions at the genetic level to help simplify the prediction of function when parts are used in combination. As a related example, ORF sets might be codon shuffled³⁶ to ‘erase’ any cryptic regulatory elements and provide a common putative mRNA secondary structure, and so on. The refining of natural parts to produce engineered biological parts may be similar to nature’s use of negative selection against promiscuous, deleterious molecular interactions within specific cell types³⁷, and is analogous to the processing of physical materials in other engineering disciplines. For comparison, microprocessors and other electronic systems are not built directly from chunks of metal and silicon found lying about the countryside.

Evolution or design of reproducing machines

Today, almost all of engineering is limited to the design and production of disposable systems. For example, our computers, mobile phones and cars are not themselves designed to directly produce the next generation of computers, phones and cars; beyond some level of failure, engineered artefacts are recycled or abandoned. To an engineer, biological systems are replicating machines that make mistakes during the replication process (that is, biological systems are reproducing machines). At present, we do not have a practical theory that supports the design of reproducing biological machines, despite great progress in understanding how natural biological systems couple and tune error detection and correction during machine replication to organism fitness (for example, see refs 38, 39). To develop such a theory, areas worth exploring may initially include reliable computing with unreliable components, error detection and correcting codes, and theory of self-replicating automata⁴⁰. If we fail to learn how to programme reproducing biological machines that we can understand, but can afford enough DNA synthesis, one fall-back option may be to engineer disposable biological systems, in which the system designs are decoupled from the constraints of direct descent and replication with error.

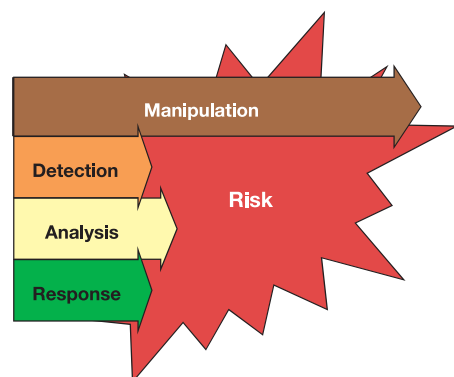


Figure 3 | Technology classes relevant to current and future biological risk. Both nature’s, and our own, ability to manipulate biological systems outpaces our ability to detect biological agents, analyse the resulting data, and respond appropriately. As a result, we are exposed to existing, emerging and engineered biological threats. Foundational technologies that affect the science and engineering of biology (for example, DNA synthesis) must be evaluated in the context of their net contribution to risk exposure and not only risk creation (see the text for a discussion).

Consequences of success

The successful development of foundational technologies based on the ideas of standardization, decoupling and abstraction would help make routine the engineering of synthetic biological systems that behave as expected. Once developed, many of these foundational technologies will take the form of ideas or information, and thus are likely to be freely distributed worldwide to the service of diverse human needs. By definition, foundational technologies would help to broadly enable many applications (that is, more than can be enumerated). The general utility of such technologies is what justifies the planning, investment and community organization needed for their development. In support of such work, what follows are a few representative examples that illustrate what should become possible within the next five years, given investment and advances in each technology class.

Standardization. Undergraduates and high school students, without prior training in biology or biological engineering, should, over a period of weeks, be able to design synthetic biological systems of their own invention comprised of several dozen pre-existing standard biological parts, order and receive the DNA encoding the system, and show it to work.

Decoupling. Systematic engineering of bacterial genomes and eukaryotic and plant chromosomes should be made possible by continued advances in DNA synthesis. Applications of genome and chromosome engineering will range from automated scientific discovery (for example, the annotation of a newly sequenced organism might be used to define a second genome that encodes only the annotated functions; synthesis of the second genome, in parts and in whole, could be used to test the accuracy of the original sequence and the sufficiency of its annotation) to biological safety (for example, orthogonal and ‘fail-fast’ codon tables that reduce the ecological impact of horizontal gene transfer) to biological defence (for example, rapid synthesis of DNA vaccines).

Abstraction. Performance characteristics and signal carriers that delimit how engineered biological parts can be combined into many-part devices and systems will be defined and may, in turn, spur the development of new measurement technologies for the characterization of cellular state⁴¹.

Biological risks, strategy and leadership

The development and distribution of foundational technologies that make the engineering of biology easier will have a direct impact on our exposure to natural and engineered biological risks. However, analysing the risk impact of such technologies is not straightforward (Fig. 3). For example, DNA synthesis has recently helped to enable the ‘resurrection’ of the 1918 influenza strain⁴², and is now widely believed capable of producing the smallpox genome from publicly available sequence information⁴³; variants of these viruses and other pathogens could be readily constructed. But, when a natural or engineered biological threat next emerges, DNA synthesis would find immediate use in accelerating our analysis of the risk (for example, rapid synthesis of pathogen ORFs codon optimized for recombinant protein expression) and our response to the risk (for example, rapid synthesis of vaccines or vaccine precursors). Meanwhile, and most importantly, the most overwhelming and certain use of DNA synthesis will be its direct acceleration of ongoing, constructive experimental research. Today, for example, a practicing experimental biologist or biological engineer can easily spend around 50% of their effort manipulating DNA just to produce the genetic material needed for an experiment; ‘instant’ DNA synthesis would provide a general twofold increase in research productivity. Thus, an attempt to ban or limit access to DNA synthesis technology, and the sequence information that defines what to synthesize, would only be guaranteed to cripple biological engineering research and hinder biomedical research. In this context, our future biological security will depend not on limits to technology and research, but rather on

the successful development of general, agile capabilities for detecting, understanding and responding to biological risks; work that is already needed given existing and emerging infectious diseases. Because technologies for engineering biology will be openly developed and widely distributed, political leadership is needed to encourage all members of society to help actively foster a worldwide community that celebrates the science of biology, and lead the overwhelmingly constructive development and application of future biological technologies⁴⁴.

- Levskaia, A. *et al.* Bacterial photography: Engineering *Escherichia coli* to see light. *Nature* doi:10.1038/nature04405 (this issue).
- Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
- Martin, V. J. *et al.* Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnol.* **21**, 796–802 (2003).
- Sprinzak, D. & Elowitz, M. B. Reconstruction of genetic circuits. *Nature* doi:10.1038/nature04335 (this issue).
- Szybalski, W. & Skalka, A. Nobel prizes and restriction enzymes. *Gene* **4**, 181–182 (1978).
- Goeddel, D. V. *et al.* Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl Acad. Sci. USA* **76**, 106–110 (1979).
- Murray, A. W. & Szostak, J. W. Construction of artificial chromosomes in yeast. *Nature* **305**, 189–193 (1983).
- Perlak, F. J. *et al.* Insect resistant cotton plants. *Biotechnology* **8**, 939–943 (1990).
- Oltvai, Z. N. & Barabasi, A. L. Systems biology. Life's complexity pyramid. *Science* **298**, 763–764 (2002).
- Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science* **295**, 1664–1669 (2002).
- Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* published online, 13 September 2005 (doi:10.1038/msb4100025).
- Weber, M. in *The Vocation Lectures* (eds Owen, D. & Strong, T. B.; translator Livingstone, R.) 1–31 (Hackett Publishing, Indianapolis, Indiana, 2004).
- Knight, T. F. Engineering novel life. *Mol. Syst. Biol.* published online, 13 September 2005 (doi:10.1038/msb4100028).
- Whitesides, G. M. The once and future nanomachine. *Sci. Am.* **285**, 78–83 (2001).
- Kirschner, M. W. The meaning of systems biology. *Cell* **20**, 503–504 (2005).
- Benner, S. A. & Sismour, A. M. Synthetic biology. *Nature Rev. Genet.* **6**, 533–543 (2005).
- Surowiecki, J. Turn of the century. *WIRED* 10.01 (<http://www.wired.com/wired/archive/10.01/standards.html>) (2002).
- International Nucleotide Sequence Database Collaboration. (<http://insdc.org/>) (2005).
- Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nature Genet.* **29**, 365–371 (2001).
- Protein Data Bank. (<http://www.rcsb.org/pdb/index.html>) (2005).
- Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), Enzyme supplement 5. *Eur. J. Biochem.* **264**, 610–650 (1999).
- Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
- Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—restriction enzymes and DNA methyltransferases. *Nucleic Acids Res.* **1**, D230–D232 (2005).
- Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
- Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behaviour in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
- Knight, T. F. DARPA BioComp Plasmid Distribution 1.00 of Standard Biobrick Components *MIT Synthetic Biology Working Group Reports* published online 22 May 2002 (doi:10.1721.1/21167).
- Graham-Rowe, D. Britain may force DNA 'barcodes' for GM food. *New Sci.* (<http://www.newscientist.com/article.ns?id=dn3377>) (2003).
- Mead, C. & Conway, L. *Introduction to VLSI Systems* (Addison-Wesley, Reading, Massachusetts, 1980).
- Canton, B. Engineering the Interface Between Cellular Chassis and Integrated Biological Systems *MIT Synthetic Biology Working Group Reports* published online 8 August 2005 (doi:10.1721.1/19813).
- Carlson, R. The pace and proliferation of biological technologies. *Bio Secur. Bioterror.* **1**, 203–214 (2003).
- Stemmer, W. P. *et al.* Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxynucleotides. *Gene* **164**, 49–53 (1995).
- Yount, B., Curtis, K. M. & Baric, R. S. Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. *J. Virol.* **74**, 10600–10611 (2000).
- Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
- Sharp, P. A. RNA interference—2001. *Genes Dev.* **15**, 485–490 (2001).
- Varmus, H. Reverse transcription. *Sci. Am.* **257**, 56–59 (1987).
- Katz, L. & Burge, C. B. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.* **13**, 2042–2051 (2003).
- Zarrinpar, A., Park, S. H. & Lim, W. A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676–680 (2003).
- Carroll, S. B. Evolution at two levels: on genes and form. *PLoS Biol.* **3**, e245 (2005).
- Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- von Neumann, J. in *Collected Works* (ed. Taub, A. H.) Vols 1–6 (Pergamon Press, New York, 1961–1963).
- Endy, D., Deese, I. & Wadey, C. Adventures in synthetic biology. *Nature* [online] (<http://www.nature.com/nature/comics/syntheticbiologycomic>) (2005); also available at (<http://mit.edu/indy/www/comic/>).
- Tumpey, T. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80 (2005).
- Venter, J. C. *Gene Synthesis Technology: State of the Science* National Science Advisory Board on Biosecurity (conference) (http://www.webconferences.com/nihnsabb/july_1_2005.html) (2005).
- Dyson, F. The darwinian interlude. *Technology Review* (<http://www.technologyreview.com/articles/05/03/issue/magaphone.asp>) (2005).
- Wolfe, S. A., Grant, R. A. & Pabo, C. O. Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry* **42**, 13401–13409 (2003).

Acknowledgements I thank T. Knight, G. Sussman and R. Rettberg for teaching me about the theory and practice of engineering, especially as it relates to biology. I also thank A. Arkin, F. Arnold, R. Baric, F. Blattner, J. Braff, R. Brent, S. Bruck, C. Bustamante, B. Canton, R. Carlson, L. Chan, A. Che, G. Church, J. Collins, C. Conboy, L. Conway, R. Davis, M. Desai, J. Doyle, E. Eisenstadt, M. Elowitz, S. Forrest, T. Gardner, S. Goldstein, J. Gritton, H. Hellinga, G. Homsy, E. Horvitz, J. Jacobson, J. Kelly, T. Kalil, J. Keasling, D. Kirkpatrick, S. Kosuri, P. Lincoln, J. Mulligan, R. Murray, R. Nagpal, R. Newton, C. Pabo, I. Phillips, P. Rabinow, R. Shetty, P. Silver, B. Smith, C. Smolke, S. Sutton, C. Tomlin, J. Way, C. Webb, R. Weiss, S. Wolfe and A. Vesilind for their contributions to the ideas discussed here. Finally, I am grateful to R. Brent and N. Kuldell for comments on an earlier manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to D.E. (indy@mit.edu).

ARTICLES

Isolation and characterization of a protochordate histocompatibility locus

Anthony W. De Tomaso¹, Spencer V. Nyholm¹, Karla J. Palmeri¹, Katherine J. Ishizuka¹, William B. Ludington¹, Katrina Mitchel¹ & Irving L. Weissman¹

Histocompatibility—the ability of an organism to distinguish its own cells and tissue from those of another—is a universal phenomenon in the Metazoa. In vertebrates, histocompatibility is a function of the immune system controlled by a highly polymorphic major histocompatibility complex (MHC), which encodes proteins that target foreign molecules for immune cell recognition. The association of the MHC and immune function suggests an evolutionary relationship between metazoan histocompatibility and the origins of vertebrate immunity. However, the MHC of vertebrates is the only functionally characterized histocompatibility system; the mechanisms underlying this process in non-vertebrates are unknown. A primitive chordate, the ascidian *Botryllus schlosseri*, also undergoes a histocompatibility reaction controlled by a highly polymorphic locus. Here we describe the isolation of a candidate gene encoding an immunoglobulin superfamily member that, by itself, predicts the outcome of histocompatibility reactions. This is the first non-vertebrate histocompatibility gene described, and may provide insights into the evolution of vertebrate adaptive immunity.

Peptide presentation by the MHC is the foundation of the vertebrate adaptive immune system. Characterized by the presence of immunoglobulins, clonal B and T cells, somatic recombination, immune memory and a number of other highly specialized processes, adaptive immunity has an incredibly complex interrelated organization that is similar in form and function in all vertebrates, from sharks to humans¹. However, as is clear by its name, the MHC was discovered because of its role in histocompatibility, and, based on this dual function, it has been long suggested that there is an evolutionary connection between histocompatibility and vertebrate immunity. Ironically, MHC-based histocompatibility is thought to be an artefact of the MHC molecules' role in the adaptive immune response coupled to their extraordinary polymorphism. Furthermore, there is no evidence of the MHC, or direct homologues of any other adaptive immune components, in more primitive extant organisms, even those as closely related as the jawless fish^{1,2}. Thus, the mechanisms underlying histocompatibility in non-vertebrates, the origins of any components of the vertebrate adaptive immune system, and the ultimate relationship of immunity and histocompatibility are completely unknown.

We are studying histocompatibility in the ascidian, *B. schlosseri*. A member of the chordate subphylum Urochordata, *Botryllus* begins life as a tadpole larva with many chordate traits, including a notochord, dorsal hollow nerve tube, and gill slits. After a 24 h motile phase, these chordate structures are lost when the tadpole settles and metamorphoses into a sessile invertebrate body plan, called an oozoid. In addition, *B. schlosseri* is a colonial organism, and this initial metamorphosis is followed by a recurring and highly coordinated budding process that eventually gives rise to a large colony of asexually derived genetically identical individuals, called zooids, united by a common vascular network (Fig. 1a, left panel).

At the periphery of the colony, the vasculature stops in small protrusions called ampullae (Fig. 1a, left panel). The ampullae are the site of a naturally occurring histocompatibility reaction initiated when two colonies asexually expand into close proximity. Once

juxtaposed, ampullae of each colony will reach out and begin an interaction (Fig. 1a, left middle panel). This will result in either fusion of the two ampullae to form a single chimaeric colony sharing a common blood supply (Fig. 1a, right middle panel), or a rejection reaction during which the interacting ampullae are destroyed, thus preventing vascular fusion (Fig. 1a, right panel). Fusion or rejection is governed by a single highly polymorphic locus called the FuHC (for fusion/histocompatibility)^{3–5}. When two colonies share one or both FuHC alleles, they will fuse; rejection occurs if no alleles are in common. The FuHC is highly polymorphic, with most populations containing tens to hundreds of alleles^{3–5}. Thus, this primitive chordate undergoes a histocompatibility reaction analogous to the vertebrate MHC-based reaction, and we aimed to uncover the molecular mechanisms underlying this process.

Positional cloning of the FuHC locus

We created conditions to raise and breed *B. schlosseri* in captivity, and also developed partially inbred lines homozygous for different FuHC alleles, which allowed us to take a forward genetic approach to identify the FuHC locus^{6–8}. As described previously⁸, mapping populations were created using FuHC-defined individuals as parents, the FuHC locus was mapped using a combination of amplified fragment length polymorphism (AFLP) and bulk segregant analysis (Fig. 1b), and a genomic walk was initiated from the linked markers using both bacterial artificial chromosome (BAC) and Fosmid genomic libraries. The physical map consisted of three contigs spanning 1.3 megabases (Mb), with one chromosomal breakpoint crossed (data not shown). Over 1 Mb of the minimal tiling path has been sequenced.

The strategy for identifying candidate FuHC gene(s) was based on the extraordinary polymorphism of the locus. Predicted genes from the genomic sequence were analysed for expression using polymerase chain reaction with reverse transcription (RT-PCR), then compared to those from the genomic clones. The genomic (BAC and Fosmid) and complementary DNA libraries were made from FuHC-defined,

¹Departments of Pathology and Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, and Department of Biology, Hopkins Marine Station, Pacific Grove, California 93950, USA.

but non-histocompatible, individuals; thus, we could concurrently survey for both expression and polymorphism of these cDNA sequences. The FuHC gene(s) would be expected to have polymorphisms with the following three characteristics: (1) absolute correlation with defined histocompatibility alleles in a fusion assay; (2) co-segregation with the same alleles in a defined cross; and, (3) extraordinary polymorphisms in colonies isolated from the wild. Finally, expression patterns should correlate spatially with the functional aspects of histocompatibility.

Isolation and characterization of a candidate FuHC gene

A Genscan (ref. 9) analysis of a sequenced contig consisting of two overlapping Fosmid clones (F531d19 and F557i23; both segregated without recombination with the FuHC; Fig. 1c) predicted a gene model encoding a transmembrane protein with an extracellular immunoglobulin domain. A full-length cDNA was isolated by rapid amplification of cDNA ends (RACE), which was 3.2 kb in length, predicting an open reading frame encoding a protein of 1007 amino acids in length, and was highly polymorphic. The candidate FuHC (cFuHC) gene encoded a type I transmembrane protein with the majority of the protein (852 amino acid residues) being extracellular, followed by a transmembrane domain and an intracellular tail of 128 amino acid residues in length.

The domain structure of the cFuHC protein is shown in Fig. 1d.

The amino terminus begins with a signal sequence, followed by an extracellular epidermal growth factor (EGF) repeat, two tandem immunoglobulin domains, and the transmembrane domain and intracellular tail. BLAST searches showed that the EGF repeat has homology to Notch and Tenascin at E values of 5×10^{-5} . The region encompassing the two immunoglobulin domains is homologous to Immunoglobulin superfamily member 4D (IGSF4D)/nectin-like 3 (NECL3) from a variety of vertebrate species ($E = 7 \times 10^{-10}$), with highest homology to chicken. No direct homologue was identified in other sequenced ascidian genomes. Three-dimensional modelling using a position-specific scoring matrix for protein fold recognition (3D-PSSM; <http://www.sbg.bio.ic.ac.uk/~3dpssm/>) suggested that the immunoglobulin domains have highest homology to the poliovirus receptor, CD155. Conserved domain searches suggest that the first immunoglobulin domain is potentially a variable, or more ancient, intermediate-type domain, and that the second is closest to a C2-type domain but is divergent and may not be easily classifiable (Supplementary Fig. 1, outlined in red). These analyses depend on the presence of conserved residues throughout the domain and conserved spacing between these residues¹⁰, however, without structural data, the true configuration remains unknown. The genomic structure of each immunoglobulin domain is also different: the first domain is encoded in three exons, while two exons code for the second domain. The intracellular tail has several

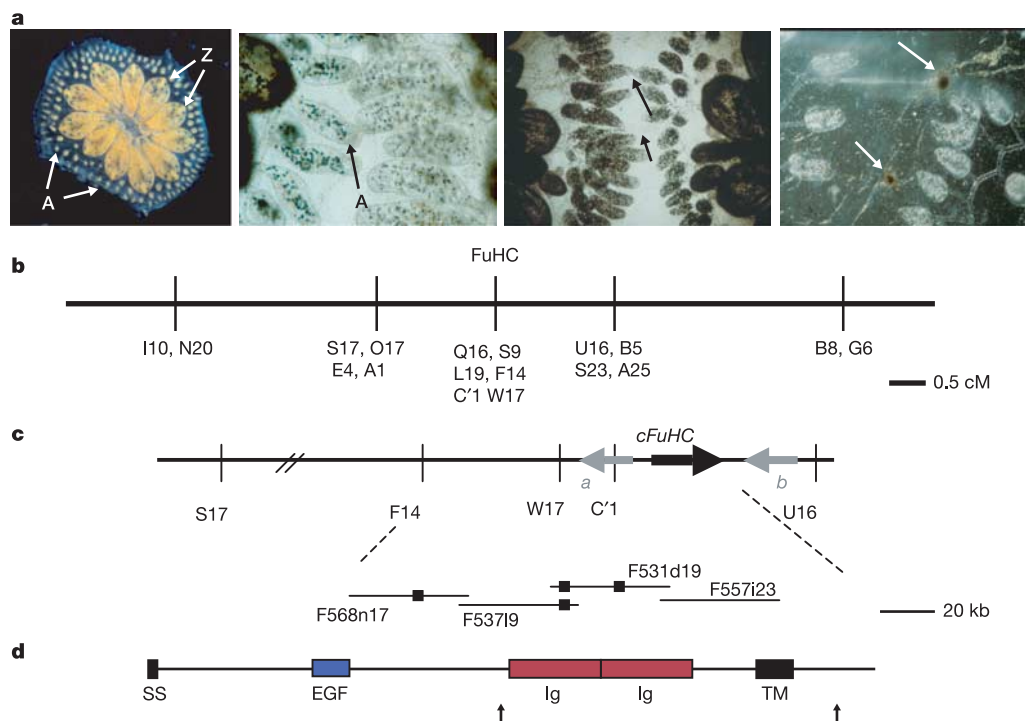


Figure 1 | Histocompatibility and positional cloning of the FuHC locus in *Botryllus schlosseri*. **a**, An adult *B. schlosseri* colony (left panel) consists of asexually derived individuals called zooids (Z) united by a common blood supply (colours are caused by pigment cells in the blood). The vasculature stops at the periphery of the colony in small protrusions called ampullae (A), where the histocompatibility reaction takes place. When two colonies grow close together, the ampullae reach out and interact (left middle panel), and this will result in one of two outcomes: either the two ampullae will fuse (right middle panel, arrows), allowing the circulation of the two colonies to interconnect, or they will reject each other (right panel). Rejection is a localized inflammatory reaction where blood cells leak from the ampullae and begin killing each other, resulting in dark spots called points of rejection (right panel, arrows) owing to melanization during the reaction. Histocompatibility takes ~24 h to occur and is controlled by a single highly polymorphic locus called the FuHC (for fusion/histocompatibility). Colonies will fuse if they share one or both alleles, and will reject if no alleles

are shared. **b**, Genetic map of the FuHC region from one backcross population. Genetic markers (AFLPs) are represented by a letter and number. In one F_2 cross (Table 1, F_2 cross AY $\times$ AY), C'1 was found to physically lie within a small 18 kb double crossover in one individual⁸; this crossover did not include the cFuHC. **c**, Physical map surrounding the candidate FuHC locus. Fosmid clones from the minimal tiling path of this region are shown on the bottom; small black boxes represent the physical location of AFLP markers, and the thick black arrow represents the location of the cFuHC locus and direction of transcription. Grey arrows show the flanking genes *a* (phospholipase A2-like) and *b* (retinoblastoma binding protein-like). **d**, Predicted structure of the FuHC protein. ss, signal sequence; EGF, epidermal growth factor repeat; Ig, immunoglobulin domain; TM, transmembrane domain. Black arrows indicate the location of two alternative splices—one that creates a secreted form of the protein and another that shortens the cytoplasmic tail (see the text).

Table 1 | cFuHC segregation in defined-cross progeny

Cross/genotype	<i>n</i>	FuHC segregation (genetic mapping)*	FuHC segregation (cFuHC allele)	Correlation
AY × AY F ₂	72	2:64:6	2:64:6	100%
AD × AD F ₂	22	1:19:3	1:19:3	100%
AC × AC F ₂	60	4:49:8	4:49:8	100%
AY × CD OC	78	16:21:19:22	16:21:19:22	100%
AA	10	10	10	100%
BB	6	6	6	100%
AB	3	3	3	100%

*FuHC genotypes were compared between the results obtained from genetic mapping versus the results from direct typing of cFuHC alleles on an individual-by-individual basis. The small number of FuHC homozygotes is discussed elsewhere²². OC, outcross.

potential phosphorylation sites, but it is unknown whether these are functional.

The cFuHC gene spans 33 kb and consists of 31 exons (Supplementary Fig. 1). In addition to the transmembrane form of the protein, which is encoded by 27 exons, there are two alternative splice variants of the cFuHC gene (Supplementary Fig. 1, bottom). First, a shorter secreted form of the protein is created by an alternative splicing event which adds three new exons after exon 14 (exons 15–17; Fig. 1d and Supplementary Fig. 1). These exons, physically located between exons 14 and 18, encode another putative low-homology EGF domain, followed by stop codons in all three reading frames and approximately 300 base pairs (bp) of 3' untranslated sequence (3'-UTR). The predicted secreted form does not include the immunoglobulin domains, and, as discussed below, is expressed both in the tadpole and in adult colonies. Second, there is a rare splice variant in the cytoplasmic domain which has been observed in 5% of the sequences covering the region. This is due to the presence of an extra 113-bp exon which is physically located between exons 29 and 31 of the full-length clone (exon 30; Fig. 1d and Supplementary Fig. 1) but is normally spliced out. When present, this exon encodes a new shorter cytoplasmic domain, and the exon ends in a stop codon. Exon 31 is still present in this alternatively spliced form, but is now included in the 3'-UTR. The shorter cytoplasmic domain adds no known motifs, and removes the predicted phosphorylation sites encoded in exon 29. This form has also been identified in both the tadpole and adult. Finally, there are also three additional splice variants with combinations of exons 15 and 16 that contain the rest of the transmembrane form of the gene (exons 18–31). However, these variants do not form open reading frames owing to frameshifts between exon boundaries when these extra exons are present in any combination. These may be non-functional transcripts made during splicing of the secreted form of the protein.

Genetics of cFuHC

To analyse segregation of the cFuHC gene, we used a fragment amplified between the exons encoding the two immunoglobulin domains, which contained parts of exons 18 and 19 and a small (250 bp) intron. As described in the Methods, this fragment was highly polymorphic, and included multiple substitutions and insertions/deletions. These polymorphisms segregated absolutely with histocompatibility (as assayed by fusion/rejection analysis or genetic mapping) for all individuals in our mapping crosses (Table 1). In addition, we correlated cFuHC polymorphisms with a number of individuals from multiple generations of our main FuHC AB × AB partially inbred lines from the last 16 years. In all cases, correlation of cFuHC polymorphisms with phenotypic and/or genetic FuHC typing was absolute.

To correlate cFuHC polymorphism with functional histocompatibility, wild-type colonies were grown in the laboratory and assayed for phenotypic fusion/rejection. We identified five robust pairs of fusing and five pairs of rejecting genotypes, including one pair of siblings which rejected each other, in addition to two pairs of fusing colonies from different geographic locations. Full-length or

extracellular polymorphic regions of the cFuHC were cloned from naive subclones of each individual and compared among genotypes. In all cases, the alleles identified by sequencing correctly predicted the outcome of histocompatibility reactions (Table 2). Thus far, we have not identified fusing genotypes that do not share an identical allele, or rejecting genotypes that do share an allele. The cFuHC polymorphisms predict the outcome of histocompatibility reactions by themselves.

These experiments also confirmed that the cFuHC is a single locus. Segregation analysis shown in Table 1 revealed no aberrant banding patterns, and the experiments shown in Table 2 on wild-type individuals never revealed more than two alleles/genotype. In addition, screening of the entire 13-fold coverage genomic library revealed only known alleles, which was confirmed by Southern blotting (data not shown).

Polymorphism and expression of cFuHC

To analyse the overall rate of polymorphism in wild-type colonies, we sequenced full-length clones from 10 wild-type individuals collected from around the Monterey Bay area. This resulted in the isolation of 18 cFuHC alleles; each individual was heterozygous. At the nucleotide level, any two alleles diverged by an average of 4% of the nucleotides. Amino acid polymorphisms in the cFuHC protein are illustrated in Fig. 2. While the alleles are quite divergent (Fig. 2a), the polymorphisms are spread throughout the protein, with no obvious highly variable regions (Fig. 2b). As shown, the majority of divergence

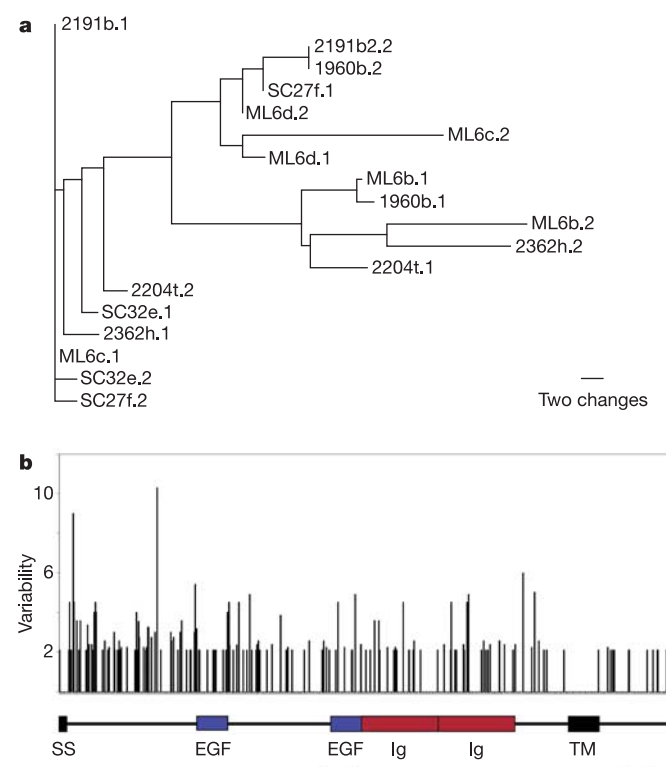


Figure 2 | Graphic representation of the overall polymorphism of the cFuHC. **a**, A phylogram of 18 cFuHC alleles isolated from wild-type animals collected in Monterey (numbers), Santa Cruz (SC) and Moss Landing (ML) marinas (see the Methods). **b**, The domain structure of the protein (bottom) is overlaid with a graphical representation of amino acid variability along the protein (top). The same 18 wild-type alleles used in **a** were aligned, and the variability is defined as the ratio of the number of different amino acids to the frequency of the most common amino acid at each position. The secreted exons and new carboxyl-terminus tail exons are also shown (underlined). There are no hypervariable regions, and the majority of the variability (that is, the shorter vertical lines) consists of single amino acid changes spread throughout the ectodomain.

between any two alleles is single amino acid changes located throughout the extracellular domain (Fig. 2b).

The final criterion for a candidate histocompatibility protein is the correct expression pattern. The FuHC gene would be expected to show expression at sites of histocompatibility in the adult, and there must also be expression in the larva; histocompatibility is functional in the oozoid, thus the effector systems must be undergoing education processes in the tadpole before metamorphosis to be able to distinguish between FuHC alleles (discussed below). RT-PCR on cDNA isolated from different tissues (Fig. 3a) revealed that both membrane-bound and secreted forms of the FuHC are expressed in the adult in both ampullae and blood, and also in the tadpole. In addition, the cFuHC does not appear to be expressed in germline tissue using this or other techniques. This corroborates previous results suggesting that the FuHC does not appear to be involved in fertilization^{11,12}.

Expression patterns were also visualized by *in situ* hybridization (Fig. 3b–i). In the oozoid, strong cFuHC expression is seen in the epithelia of the ampullae, as well as in a subset of blood cells. Expression in the tadpole is restricted to a region at the anterior end of the larvae, outlining the nascent ampullae. Expression in isolated ripe eggs is not seen. In summary, the cFuHC is expressed strongly in tissues intimately associated with histocompatibility.

Discussion

We have isolated and characterized a highly polymorphic candidate

fusion/histocompatibility (cFuHC) gene from the primitive chordate *Botryllus schlosseri*, which encodes an immunoglobulin superfamily member. The cFuHC segregates absolutely with histocompatibility in all mapping crosses, by itself predicts the outcome of interactions between wild-type colonies, and is expressed in tissues directly associated with the natural transplantation reaction: by every criterion this candidate is directly responsible for controlling histocompatibility. Other genes with these characteristics have not been identified in the FuHC locus, and, together with the additional data presented here, this is strong evidence that the cFuHC is the first metazoan histocompatibility ligand identified outside of the vertebrate lineage.

The ability to discriminate between multiple ligands is the universal principle shared by immunity and histocompatibility, leading Burnet to hypothesize that the origins of the most complex recognition system in the Metazoa, the vertebrate adaptive immune system, would be found in the histocompatibility systems of more primitive organisms¹³. Does the identification of the cFuHC in a primitive chordate confirm this hypothesis? As a type I transmembrane protein with multiple extracellular immunoglobulin domains, the cFuHC is certainly structurally similar to the vertebrate MHC. However, it is clearly not a direct homologue; for example, the immunoglobulin domains do not correspond to the C1 type found in the MHC. Alternatively, it has been suggested that vertebrate genes with more ancient immunoglobulin domains, like human IGSF4 (with which the cFuHC has high homology), may be descendents of the first antigen receptors^{2,14}. This is also based on observations of the

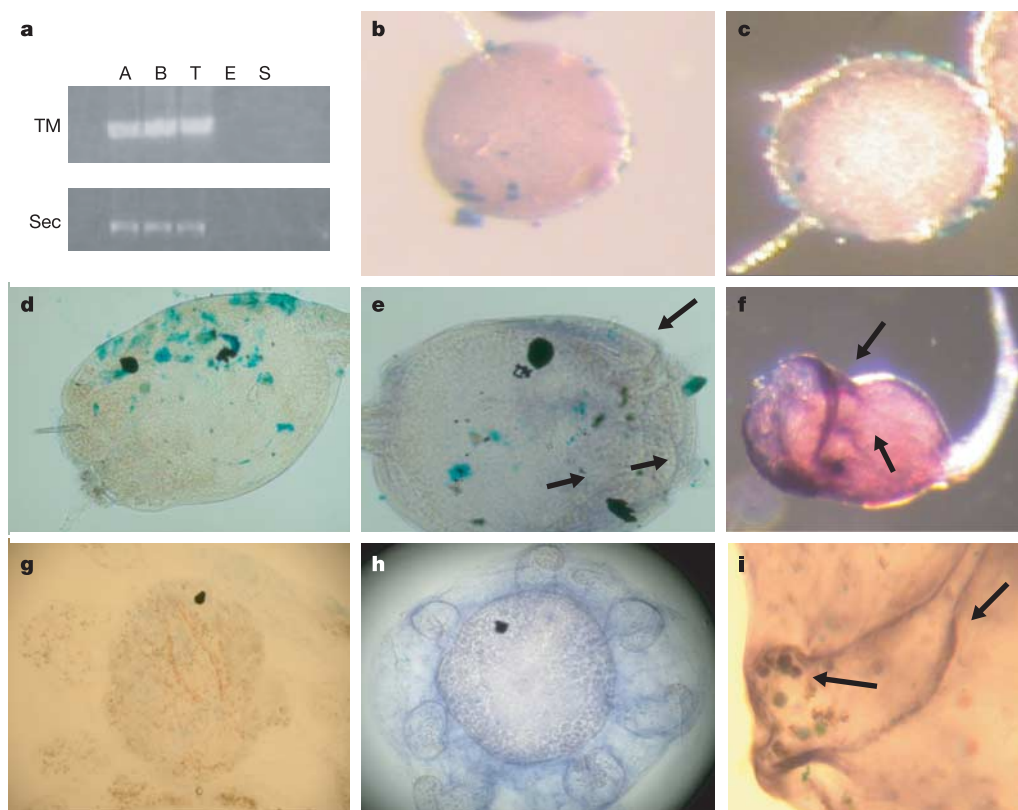


Figure 3 | Expression analysis of the cFuHC. **a**, RT-PCR of cDNA isolated from ampullae (A), blood (B), tadpoles (T), eggs (E) and sperm (S). Top panel, 600-bp fragment including the transmembrane domain (TM); bottom panel, 350-bp fragment including the secreted 3'-UTR (Sec). Both primer sets straddle a large intron, preventing amplification from genomic DNA. **b–i**, Localization of cFuHC expression by *in situ* hybridization in eggs, tadpoles and oozoids (described in the Methods). Panels **b**, **d** and **g** are sense controls. Eggs (**c**) were isolated immediately before fertilization but showed no cFuHC expression when compared with controls. In the tadpole,

expression was restricted to the anterior of the larvae, and outlines the nascent ampullae (**e**, arrows). A larvae in the middle of metamorphosis (**f**) showed strong expression at the tips of the extending ampullae (arrows). Expression in the adult body was seen in the epithelia of the ampullae (**h**). A close-up of the ampullae is shown in **i**, with expression seen in both a subset of blood cells and the epithelium of the ampullae and blood vessel. Light blue staining is Vector Blue, dark purple is BCIP/NBT. The punctate light blue staining seen in **b–e** is a result of Vector Blue precipitation, which often occurs during colour development.

Table 2 | Phenotypic outcome compared with FuHC genotype

Colony 1	Colony 2	Relationship	Histocompatibility outcome*	cFuHC shared†
1960b (het)‡	1960c (het)	Sibs	Fuse	1
2204t (het)	2204p (het)	Sibs	Fuse	1
2190b (het)	2190c (het)	Sibs	Fuse	1
SC27f (het)	HM9aY (het)	None	Fuse	1
SC32e (het)	HM9aY (het)	None	Fuse	1
ML6c (het)	2191b (het)	None	Reject	0
2191b (het)	ML6b (het)	None	Reject	0
SC27f (het)	SC32e (het)	None	Reject	0
HM9aY (het)	2362h (het)	None	Reject	0
ML6c (het)	ML6d (het)	Sibs	Reject	0

* Individuals were placed together and fusion or rejection visibly observed. Instances where unrelated colonies underwent fusion owing to a shared cFuHC allele, or where related colonies underwent rejection owing to no shared alleles, are shown in bold.

† FuHC alleles were genotyped from naive subclones and compared (described in the Methods).

‡ Zygosity at the cFuHC locus is shown (het, heterozygous).

organization of these genes in mammalian genomes², which may also be shared with the FuHC (data not shown). Further structural and genetic analysis of the cFuHC may help elucidate the evolutionary origins of the different immunoglobulin domains, and ultimately the molecules in adaptive immunity.

From a functional standpoint, there is another possible relationship between histocompatibility in *Botryllus* and the origins of vertebrate immunity. In *Botryllus*, individuals sharing only a single allele are compatible. It seems unlikely that hundreds of FuHC alleles could evolve to clearly distinguish themselves through only homo-typic interactions, suggesting that the effector systems in *Botryllus* work by missing-self recognition, a strategy first identified in vertebrate natural killer (NK) cells¹⁵. We hypothesize that a population of effector cells are educated to both FuHC alleles, and that recognition of just one allele blocks a default rejection reaction. Histocompatibility in *Botryllus* evolutionarily links an innate function of vertebrate NK cells, scanning for altered MHC expression, to a highly polymorphic immunoglobulin superfamily ligand. Thus, missing-self recognition and natural killing effector systems may have originated in histocompatibility systems of non-vertebrates. In fact, the genome of the related ascidian *Ciona intestinalis* contains a number of genes highly homologous to those involved in natural killing in the vertebrates¹⁶. A principal goal of future research will be to identify the cells, molecules and mechanisms underlying the function and education of this effector system. Notably, we have already found what appears to be a putative polymorphic receptor encoded within 200 kb of the cFuHC. However, these polymorphisms play no role in histocompatibility outcomes (S.V.N., E. Passegue, I.L.W., W.B.L., A. Voskoboynik, K.M. and A.W.D., manuscript in preparation).

Vertebrate MHC-based histocompatibility is thought to be an unintended consequence of peptide presentation and polymorphism, but histocompatibility and, more importantly, extraordinary polymorphism have a well-defined function in *B. schlosseri*. After compatible colonies fuse, blood-borne germline or totipotent stem cells transfer between colonies and have the ability to expand and differentiate in the newly arising, asexually derived individuals of the vascular partner. This can often result in a situation where only one of the genotypes is represented in the gametic output of the fused individuals, a situation that can remain constant for life^{17–19}. FuHC polymorphism functions to restrict vascular fusion, and the possibility of this germline parasitism, to kin. This stem cell parasitism is thought to be widespread, and may be the selective force driving the evolution of polymorphic histocompatibility systems in other phyla^{20,21}. Moreover, this movement of stem cells in *Botryllus* is reminiscent of the transfer of haematopoietic stem cells between fetal mammals sharing a common placenta, the basis of natural tolerance²². These same studies did not identify germline chimaeras

despite the fact that mammalian germline precursors are also migratory during development^{22,23}. Thus, it may be that histocompatibility is still one function of vertebrate immunity, providing both a historical origin and a contemporary selective force to explain extraordinary MHC polymorphism. Alternatively, the FuHC may represent an undiscovered immunoglobulin-based system of allo-recognition that still exists in the vertebrates; for example, NK cells have recently been found to interact with a non-MHC ligand through missing-self recognition^{24,25}. Furthermore, CD155 (a nectin family member with which the cFuHC shares structural homology) is a ligand for the NK activating receptor CD226 (ref. 26). In any case, the cFuHC is the first non-vertebrate histocompatibility ligand to be identified, and as an immunoglobulin superfamily member provides the first structural link between invertebrate histocompatibility and vertebrate adaptive immunity.

METHODS

Mariculture and collection of *B. schlosseri*. Conditions for raising and crossing *B. schlosseri* in the laboratory, and techniques for assaying phenotypic histocompatibility, have been comprehensively described elsewhere^{6,8}. For fusing or rejecting pair analysis, pregnant colonies were collected from the wild, and then hatched and reared in the laboratory. One naive subclone was removed and DNA and RNA isolated; another subclone was used in a fusion/rejection assay^{3–5}. Animals are named according to their collection site. Those starting with a number are from the Monterey Marina, California; other collection sites are Half Moon Bay (HM), Santa Cruz (SC), and Moss Landing (ML), California. This is followed by a lowercase letter which indicates an F₁ individual; thus, individuals 1960b and 1960c are progeny from the same wild-type colony.

Isolation and analysis of a candidate FuHC gene. Genetic and physical mapping of the FuHC locus, sequencing, DNA and RNA isolation, and cDNA synthesis have been comprehensively described elsewhere^{6,8}. Enzymes were from NEB. All chemicals were purchased from Fisher. Oligonucleotides were synthesized by Operon. Kits were used for specific processes as described below. BAC and Fosmid sequences were analysed using BLAST (ref. 27) and GENSCAN (ref. 9). Putative expressed genes were analysed for topology and functional domains using the HMMTOP (ref. 28) transmembrane predictor and ELM (ref. 29). Sequences which caught our interest were then isolated by RACE (ref. 30) and RT-PCR using the SMART II RACE kit as described (Clontech). Initial primers for the cFuHC were in two regions. STS (sequence tagged site) 1 covered the immunoglobulin domain: sense, 5'-GTATGGGACAACACAGGAAATTCTAC-3'; antisense, 5'-GTGACGTTTGTAGTCCATAGGATATCAG-3'. STS 2 was upstream of this region: sense, 5'-TACTATTGAGTGTATGAACGGTGATGT-3'; antisense, 5'-TTCTATCGCCCTATATAGTTTGTAA-3'. These same primers were used to isolate the entire gene by RACE. The entire gene was amplified with the following primers: the sense primer, 5'-AAGCATG AATGGGTTTCGCGATTTC-3' (encompassing the start codon); and one of two antisense primers for either the secreted form, 5'-TCGTTATGCTGTATT CATTCAA-3', or the membrane-bound form, 5'-AAGCTCTTTCAGAGC TACTATCTTCA-3' (each located near the respective stop codons of each form of the messenger RNA). For the membrane-bound form, we usually amplified in two overlapping ~1.5 kb fragments from primers designed over the exon 14–exon 18 junction. The forward primer was 5'-TACGTTGATTGG AACTGTCGACTTGAAGTA-3' for amplifying the 3' end, and the reverse primer was 5'-TACTTCAAGTCGACAGTTCCAATCAACGTA-3' for amplifying the 5' end, used in conjunction with the primers described above. This second strategy was used as there are a number of alternative splice variants that encode non-functional transcripts in exons 15–17. None of these variants form open reading frames, owing to frameshifts in the exon splice boundaries (data not shown). As these are near the secreted alternative splice site (after exon 16), we hypothesize that they are non-functional transcripts made during mRNA processing.

Segregation and expression analysis. The majority of segregation analysis was done using STS 1 (described above) as the forward and reverse primers were each on different exons and split by a small (250 bp) intron. This intron was highly polymorphic, and the nature of these polymorphisms allowed a PCR–restriction fragment length polymorphism (PCR–RFLP)-based approach to distinguish between all of the FuHC alleles in our mapping populations. The FuHC A, B and Y alleles could be differentiated by *EcoRI*, *HindIII* and *HindII* polymorphisms, and these techniques have been previously described⁸. Segregation was confirmed by sequencing the entire clone from recombinant individuals in each cross (data not shown). To detect expression by RT-PCR, we used STS 1 and 2, as well as two other primer sets tailored for the membrane-bound or secreted form.

The membrane-bound form RT-PCR primer set was: STS 1 sense, shown above; antisense, 5'-GTACCTCAAGTACCACAGCCCCAAT-3'. The secreted form was: sense, 5'-TGCAGCTTGGAAGTTTTC-3'; antisense, 5'-TCGTTATGCTGATTCAATCA-3'. Protein polymorphisms and phylograms were prepared using ClustalW (ref. 31).

***In situ* hybridization.** *In situ* probes, amplified using the primers described above, were subcloned into pGEM-T vector (Promega), sequenced to determine orientation, then re-amplified with T7 and sp6 primers. After purification using a PCR cleanup kit (Qiagen), *in situ* hybridization probes were made using T7 and sp6 RNA polymerase according to the manufacturer's instructions (Boehringer). The whole mount *in situ* hybridization protocol used in this study was carried out according to references 32 and 33, on laboratory-reared individuals. Signal detection was carried out using either 5-bromo-4-chloro-3'-indolylphosphate/nitroblue tetrazolium (BCIP/NBT) or a Vector Blue alkaline phosphatase substrate kit (Vector Laboratories) with levamisole. After colour development, samples were post-fixed in 4% (w/v) paraformaldehyde in PBS (20 min at 4 °C), washed in PBS, mounted on slides, and visualized on an Olympus BH-2 microscope.

Received 17 June; accepted 18 August 2005.

- The membrane-bound form RT-PCR primer set was: STS 1 sense, shown above; antisense, 5'-GTACCTCAAGTACCACAGCCCCCAAT-3'. The secreted form was: sense, 5'-TGCAGCTTGGAAAGTTTGTG-3'; antisense, 5'-TCGTTATGC TGTATTCATTCAA-3'. Protein polymorphisms and phylograms were prepared using ClustalW (ref. 31).
- In situ* hybridization.** *In situ* probes, amplified using the primers described above, were subcloned into pGEM-T vector (Promega), sequenced to determine orientation, then re-amplified with T7 and sp6 primers. After purification using a PCR cleanup kit (Qiagen), *in situ* hybridization probes were made using T7 and sp6 RNA polymerase according to the manufacturer's instructions (Boehringer). The whole mount *in situ* hybridization protocol used in this study was carried out according to references 32 and 33, on laboratory-reared individuals. Signal detection was carried out using either 5-bromo-4-chloro-3'-indolylphosphate/nitroblue tetrazolium (BCIP/NBT) or a Vector Blue alkaline phosphatase substrate kit (Vector Laboratories) with levamisole. After colour development, samples were post-fixed in 4% (w/v) paraformaldehyde in PBS (20 min at 4°C), washed in PBS, mounted on slides, and visualized on an Olympus BH-2 microscope.
- Received 17 June; accepted 18 August 2005.**
1. Flajnik, M. F., Miller, K. & Dupasquier, L. in *Fundamental Immunology* 5th edn (ed. Paul, W. E.) 519–570 (Lippincott, Williams and Wilkins, Philadelphia, 2003).
 2. Dupasquier, L., Zucchetti, I. & De Santis, R. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* **198**, 233–248 (2004).
 3. Oka, H. & Watanabe, H. Problems of colony specificity in compound ascidians. *Bull. Mar. Biol. Stn Asamushi* **10**, 153–155 (1960).
 4. Sabbadin, A. Le basi genetiche dell capacita di fusione fra colonie in *Botryllus schlosseri* (Ascidacea). *Atti Accad. Naz. Lincei Rend.* **32**, 1031–1035 (1962).
 5. Scofield, V. L., Schlumpberger, J. M., West, L. A. & Weissman, I. L. Protochordate allorecognition is controlled by an MHC-like gene system. *Nature* **295**, 499–502 (1982).
 6. Boyd, H. C., Brown, S. K., Harp, J. A. & Weissman, I. L. Growth and sexual maturation of laboratory-cultured Monterey *Botryllus schlosseri*. *Biol. Bull.* **170**, 91–109 (1986).
 7. De Tomaso, A. W., Saito, Y., Ishizuka, K. J., Palmeri, K. J. & Weissman, I. L. Mapping the genome of a model protochordate. I. A low resolution genetic map encompassing the fusion/histocompatibility (Fu/Hc) locus of *Botryllus schlosseri*. *Genetics* **149**, 277–287 (1998).
 8. De Tomaso, A. W. & Weissman, I. L. Initial characterization of a protochordate histocompatibility locus. *Immunogenetics* **55**, 480–490 (2003).
 9. Burge, C. B. & Karlin, S. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* **8**, 346–354 (1998).
 10. Williams, A. F. & Barclay, A. N. The immunoglobulin superfamily—domains for surface recognition. *Annu. Rev. Immunol.* **6**, 381–405 (1988).
 11. Grosberg, R. K. & Hart, M. W. Mate selection and the evolution of highly polymorphic self/non-self recognition genes. *Science* **289**, 2111–2114 (2000).
 12. De Tomaso, A. W. & Weissman, I. L. Evolution of a protochordate allorecognition locus. *Science* **303**, 977 (2004).
 13. Burnet, F. M. Self-recognition in colonial marine forms and flowering plants in relation to the evolution of immunity. *Nature* **232**, 123–126 (1971).
 14. Cannon, J. P., Haire, R. N., Rast, J. P. & Litman, G. W. The phylogenetic origins of the antigen-binding receptors and somatic diversification mechanisms. *Immunol. Rev.* **200**, 200–212 (2004).
 15. Kärre, K., Ljunggren, H. G., Piontek, G. & Kiessling, R. Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. *Nature* **319**, 675–678 (1986).
 16. Azumi, K. *et al.* Genomic analysis of immunity in a basal chordate and evolution of the vertebrate immune system: waiting for Godot. *Immunogenetics* **55**, 570–581 (2003).
 17. Izzard, C. S. Migration of germ cells through successive generations of pallial buds in *Botryllus schlosseri*. *Biol. Bull.* **135**, 424–431 (1968).
 18. Sabbadin, A. & Zaniolo, G. Sexual differentiation and germ cell transfer in the colonial ascidian *Botryllus schlosseri*. *J. Exp. Zool.* **207**, 279–301 (1979).
 19. Stoner, D. S., Rinkevich, B. & Weissman, I. L. Heritable germ and somatic cell lineage competitions in chimeric colonial protochordates. *Proc. Natl Acad. Sci. USA* **96**, 9148–9153 (1999).
 20. Buss, L. W. Somatic cell parasitism and the evolution of somatic tissue. *Proc. Natl Acad. Sci. USA* **79**, 5337–5341 (1982).
 21. Buss, L. W. *The Evolution of Individuality* (Princeton Univ. Press, New Jersey, 1987).
 22. Owen, R. D. Immunologic consequence of vascular anastomoses between twins. *Science* **210**, 400–401 (1945).
 23. Mintz, B. & Russel, E. S. Gene-induced embryological modifications of primordial germ cells in the mouse. *J. Exp. Zool.* **134**, 207–237 (1957).
 24. Carlyle, J. R. *et al.* Missing self-recognition of Oci/Clr-b by inhibitory NKR-P1 natural killer cell receptors. *Proc. Natl Acad. Sci. USA* **101**, 3527–3532 (2004).
 25. Iizuka, K., Naidenko, O. V., Plougastel, B. F., Fremont, D. H. & Yokoyama, W. M. Genetically linked C-type lectin-related ligands for the NKR1 family of natural killer cell receptors. *Nature Immunol.* **4**, 801–807 (2003).
 26. Bottino, C. *et al.* Identification of PVR (CD155) and Nectin-2 (CD112) as cell surface ligands for the human DNAM-1 (CD226) activating molecule. *J. Exp. Med.* **198**, 557–567 (2003).
 27. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
 28. Tusnády, G. E. & Simon, I. Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. *J. Mol. Biol.* **283**, 489–506 (1998).
 29. Puntervoll, P. *et al.* ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**, 3625–3630 (2003).
 30. Frohman, M. A., Dush, M. K. & Martin, G. R. Rapid production of full-length cDNAs from rare transcripts; amplification using a single gene-specific oligonucleotide primer. *Proc. Natl Acad. Sci. USA* **85**, 8998–9002 (1987).
 31. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
 32. Davidson, B. & Swalla, B. J. A molecular analysis of ascidian metamorphosis reveals activation of an innate immune response. *Development* **129**, 4739–4751 (2002).
 33. Bassham, S. & Postlethwait, J. *Brachyury (T)* expression in embryos of a larvacean urochordate, *Oikopleura dioica*, and the ancestral role of *T*. *Dev. Biol.* **220**, 322–332 (2000).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank C. Amemiya, D. Rokhsar, R. Davis, A. Southwick, M. Miranda, D. Ransom, J. Cannon, G. Litman, R. Haire, A. Voskoboinik, J. Wallace and E. Johnson. This study was supported by grants from the National Institutes of Health (a Trans-BAC Sequencing grant) and the Community Sequencing Program at the Department of Energy Joint Genome Institute.
- Author Information** Sequence information has been deposited at the NCBI with the accession numbers DQ160291 (for the membrane-bound form) and DQ160292 (for the secreted form). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.W.D. (tdet@stanford.edu).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Amemiya, D. Rokhsar, R. Davis, A. Southwick, M. Miranda, D. Ransom, J. Cannon, G. Litman, R. Haire, A. Voskoboinyk, J. Wallace and E. Johnson. This study was supported by grants from the National Institutes of Health (a Trans-BAC Sequencing grant) and the Community Sequencing Program at the Department of Energy Joint Genome Institute.

Author Information Sequence information has been deposited at the NCBI with the accession numbers DQ160291 (for the membrane-bound form) and DQ160292 (for the secreted form). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.W.D. (tdet@stanford.edu).

ARTICLES

Direct observation of base-pair stepping by RNA polymerase

Elio A. Abbondanzieri^{1*}, William J. Greenleaf^{1*}, Joshua W. Shaevitz^{2†}, Robert Landick⁴ & Steven M. Block^{1,3}

During transcription, RNA polymerase (RNAP) moves processively along a DNA template, creating a complementary RNA. Here we present the development of an ultra-stable optical trapping system with ångström-level resolution, which we used to monitor transcriptional elongation by single molecules of *Escherichia coli* RNAP. Records showed discrete steps averaging 3.7 ± 0.6 Å, a distance equivalent to the mean rise per base found in B-DNA. By combining our results with quantitative gel analysis, we conclude that RNAP advances along DNA by a single base pair per nucleotide addition to the nascent RNA. We also determined the force-velocity relationship for transcription at both saturating and sub-saturating nucleotide concentrations; fits to these data returned a characteristic distance parameter equivalent to one base pair. Global fits were inconsistent with a model for movement incorporating a power stroke tightly coupled to pyrophosphate release, but consistent with a brownian ratchet model incorporating a secondary NTP binding site.

Processive molecular motors tend to move in discrete steps¹. Recent advances in single-molecule techniques have made it possible to observe such steps directly at length scales of a few nanometres or greater. The ability to detect individual catalytic turnovers, as monitored through motor displacement, while simultaneously controlling the force, substrate concentration, temperature or other parameters, provides a means to probe the mechanisms responsible for motility. Single-molecule measurements of stepping have supplied fresh insight into the mechanisms responsible for motion in motor proteins such as myosin, kinesin, dynein and the F_1 -ATPase^{2–9}. A number of processive nucleic acid-based enzymes, such as lambda exonuclease^{10,11}, RecBCD helicase^{12–14} and RNAP^{15–19}, have also been studied successfully by single-molecule methods, but the comparatively small size of their steps has been experimentally inaccessible up to this point. Movements through a single base pair along double-stranded DNA correspond to a displacement of just ~ 3.4 Å (ref. 20), which is more than 20-fold smaller than the 8-nm kinesin step⁴ and sevenfold smaller than the 2–3-nm resolution limit attained in most previous work^{2,3,14,19,21}.

During transcription, *E. coli* RNAP translocates along DNA while following its helical pitch²², adding ribonucleoside triphosphates (NTPs) successively to the growing RNA. The basic reaction cycle consists of binding the appropriate NTP, incorporation of the associated nucleoside monophosphate into the RNA, and release of pyrophosphate. In addition to following the main reaction pathway, RNAP can reversibly enter any of several off-pathway paused states. For example, RNAP may backtrack by several bases along DNA, displacing the RNA 3' end from the catalytic centre and temporarily inactivating the enzyme^{23–25}. Single-molecule studies have shown that backtracking pauses are enhanced under hindering loads and can be triggered by misincorporation of non-complementary NTPs¹⁹. In a separate class of pauses, hairpin structures formed in the nascent RNA can induce pausing via a distinct mechanism²⁶. A third class of pauses occurs commonly and is independent of backtracking¹⁸ or hairpin formation. Regardless of the cause, paused states complicate the interpretation of biochemical studies of RNAP

elongation because these kinetic measurements convolve on- and off-pathway events into overall rates. Single-molecule techniques avoid ensemble averaging and permit, in principle, observations that may distinguish between elongation and off-pathway states. The resolution of individual enzymatic turnovers would be especially helpful in unravelling the behaviour of complex reaction cycles for enzymes such as RNAP.

The mechanism that leads to translocation during transcriptional elongation continues to be debated^{27–32}, and at least two classes of models have been proposed. The first class postulates that a power stroke tightly coupled to pyrophosphate release drives motion²⁹. In the second class, reversible diffusion of the enzyme along the DNA template between its pre- and post-translocated states is directionally rectified through the binding of the incoming NTP, in a brownian ratchet mechanism^{30–32}. By detecting individual translocation events during transcription and characterizing the force and nucleotide sensitivity of the corresponding motions, one can distinguish between the different classes of mechanism.

Construction of an ultra-stable optical trap

Because the distance spanned by a base pair is so small, it was necessary to construct a stable optical trapping system capable of ångström-level resolution. Sources of noise that hamper optical measurements include drift of the microscope stage (or other nominally stationary components), pointing fluctuations leading to relative motions of the laser beams used for trapping and position detection, and brownian motions of the trapped bead itself. To minimize the noise associated with stage motions, we used a dual-trap 'dumbbell' arrangement, described previously¹⁹. In this experimental geometry (Fig. 1a), all components of the assay are optically levitated above the coverglass surface and thereby decoupled from stage drift. A stalled transcription complex¹⁸ containing a biotin tag on the carboxy terminus of the β' -subunit was specifically attached via an avidin linkage to the surface of a 600-nm diameter polystyrene bead. Depending on the desired direction of applied load, either the transcriptionally upstream or downstream end of the DNA was then

¹Department of Applied Physics, ²Department of Physics, ³Department of Biological Sciences, Stanford University, Stanford, California 94305, USA. ⁴Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706, USA. [†]Present address: Department of Integrative Biology, University of California, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

bound via a digoxigenin antibody linkage to a 700-nm diameter bead, forming a bead–DNA–RNAP–bead dumbbell. Dumbbells were suspended $\sim 1\ \mu\text{m}$ above the microscope coverglass by two independently steered traps, T_{weak} and T_{strong} .

To isolate the detection and trapping beams from the effects of random air currents, which introduce density fluctuations that perturb the positional stability of laser beams, we enclosed all optical elements external to the microscope in a sealed box filled with helium gas at atmospheric pressure. Because the refractive index of helium is closer to unity than that of air ($n_{\text{He}} = 1.000036$ compared with $n_{\text{air}} = 1.000293$), density fluctuations introduce smaller deflections.

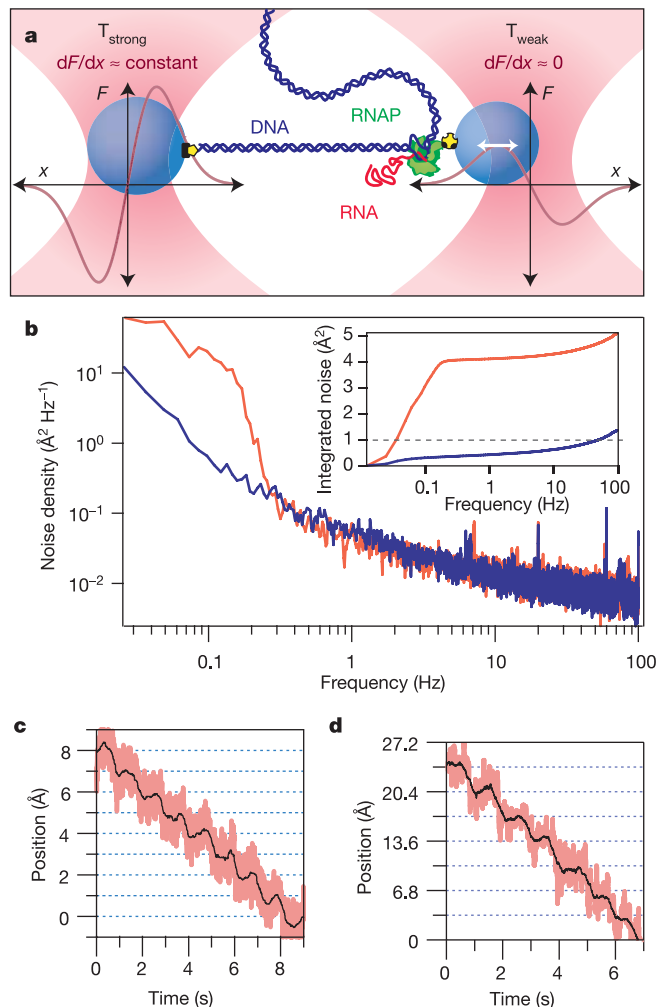


Figure 1 | Experimental set-up, passive force clamp and sensitivity of the RNAP dumbbell assay. **a**, Cartoon of the dumbbell geometry with schematic force versus position curves (dark red) shown for both trap beams (not drawn to scale). A single, transcriptionally active molecule of RNAP (green) is attached to a bead (blue) held in trap T_{weak} (pink, right) and tethered via the upstream DNA (dark blue) to a larger bead held in trap T_{strong} (pink, left). The right bead is maintained at a position near the peak of the force-extension curve of T_{weak} , where trap stiffness vanishes (white arrow), creating a force clamp (trap stiffness $k = dF/dx$). During elongation, the DNA tether lengthens and the beads move apart. Owing to the force clamp arrangement, only the right bead moves: displacement is measured for this bead. **b**, Power spectrum acquired for a stiffly trapped bead with external optics under air (red) or helium (blue). Inset: integrated noise spectra for air (red) and helium (blue) showing a tenfold reduction in power. **c**, Steps resolved for a stiffly trapped bead moved in 1-Å increments at 1 Hz. Data were median filtered with a 5-ms (pink) and 500-ms (black) window. **d**, Steps resolved for a bead–DNA–bead dumbbell held at 27 pN of tension, produced by moving T_{strong} in 3.4-Å increments at 1 Hz and measuring the corresponding displacements in T_{weak} .

Using helium, we realized a tenfold reduction in the noise spectral density at 0.1 Hz for a stiffly trapped, 700-nm diameter bead ($k = 1.9\ \text{pN nm}^{-1}$), and the integrated system noise power remained below $\sim 1\ \text{\AA}$ over the bandwidth of interest (Fig. 1b). To illustrate the resolution achieved, we moved a trapped bead in increments of 1 Å at 1-s intervals by displacing T_{strong} with an acousto-optic deflector (AOD)³³. Steps were clearly resolved, with signal-to-noise ratio of ~ 1 over a 100-Hz bandwidth (Fig. 1c).

Finally, we implemented a recently developed method to maintain constant force on trapped beads using an all-optical arrangement, without the need for computer feedback³⁴. Such a passive force clamp eliminates artefacts associated with feedback loops and provides exceedingly high bandwidth. To create this clamp, the bead in T_{weak} was maintained in a $\sim 50\text{-nm}$ region near the maximum of the force-extension curve, where the force is independent of bead position. Force clamps eliminate the need for elastic corrections due to either the compliance of T_{strong} or the stretching of the DNA along with its associated linkages, so that all molecular displacements are registered in the motion of the bead in T_{weak} . To demonstrate the resolution achieved in our set-up, a dumbbell consisting of beads connected by a DNA tether (but no RNAP enzyme) was stepped in increments of 3.4 Å at 1 Hz (Fig. 1d).

RNA polymerase takes single-base-pair steps

To resolve individual translocation events, we required that RNAP transcribe slowly enough to time-average to the ångström level over positional uncertainties caused by brownian motions, but quickly enough so that long-term drift did not obscure motion. Because RNAP has different average rates of addition for the four species of nucleotide³⁵, we determined by gel analysis the concentration ratios at which each species becomes equally rate limiting for elongation on our template. Unless otherwise noted, our experiments were conducted at $[\text{NTP}]_{\text{eq}} = 10\ \mu\text{M}$ GTP, $10\ \mu\text{M}$ UTP, $5\ \mu\text{M}$ ATP and $2.5\ \mu\text{M}$ CTP, concentrations that produce a mean elongation rate of ~ 1 base pair (bp) s^{-1} under our conditions (see Supplementary Fig. S1).

In single-molecule records of transcription by RNAP selected for their low noise and drift, we observed clear, stepwise advancements. Figure 2a shows six representative traces obtained under 18 pN of assisting load. Although dwells at some expected positions (Fig. 2a, dotted lines) were missed or skipped, steps were uniform in size, corresponding to nearly integral multiples of a common spacing. To estimate this fundamental spacing, we performed a periodogram analysis³⁶. The position histograms for 37 segments derived from transcription records for 28 individual RNAP molecules were computed and the autocorrelation function calculated for each of these³⁷. These autocorrelations were combined into a global average that displays a series of peaks near multiples of the mean spacing (Fig. 2b), with the first and strongest peak at $3.4 \pm 0.8\ \text{\AA}$. The power spectral density of this function measures the corresponding spatial frequencies and displays a prominent peak at the inverse of $3.7 \pm 0.6\ \text{\AA}$ (Fig. 2c). This distance is consistent with the crystallographic spacing between neighbouring base pairs in B-DNA³⁸ ($3.4 \pm 0.5\ \text{\AA}$). Although the foregoing analysis was performed on selected traces, a fully automated procedure was also conducted on a continuous, $\sim 300\text{-bp}$ record of elongation, and returned a similar spacing of $3.7 \pm 1.5\ \text{\AA}$ (Supplementary Fig. S2).

In our records, RNAP does not dwell at every base-pair position along the template. Were certain bases skipped altogether, or merely missed due to finite time resolution? Because RNAP is linked to the bead via the C terminus of its β' -subunit, skipped steps might be explained, in principle, by relative motions of this point of attachment with respect to the catalytic core of the enzyme, allowing the active site to undergo one or more rounds of nucleotide addition before translocation of the attachment point in a single jump, constituting a form of ‘inchworming’ movement. Alternatively, if the upstream DNA exists in a ‘scrunched’ state within the enzyme after templating the production of RNA, as proposed to occur during

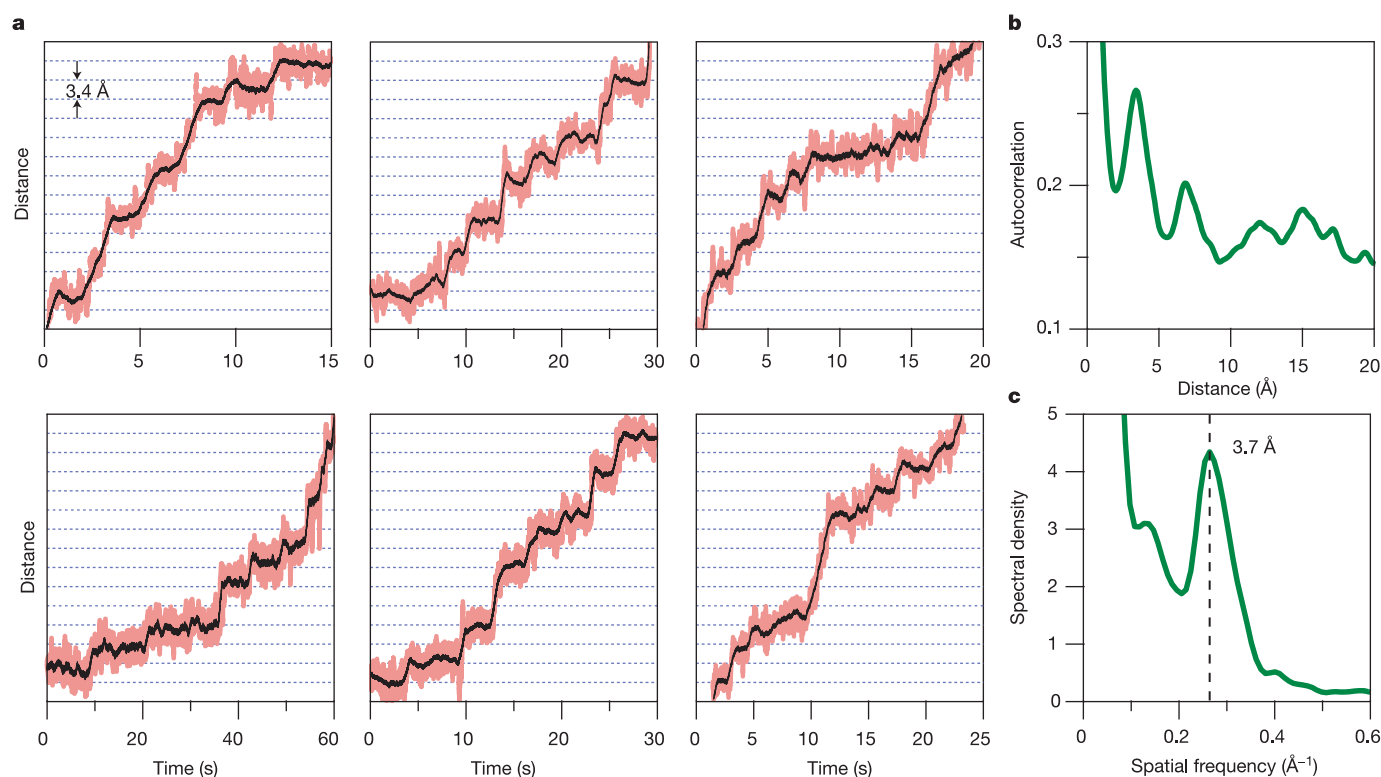


Figure 2 | RNAP moves in discrete steps. **a**, Representative records for single molecules of RNAP transcribing at $[NTP]_{eq}$ under 18 pN of assisting load, median-filtered at 50 ms (pink) and 750 ms (black). Horizontal lines (dotted) are spaced at 3.4-Å intervals. **b**, The average autocorrelation

function derived from position histograms ($N = 37$) exhibits periodicity at multiples of the step size. **c**, The power spectrum of **b** shows a peak at the dominant spatial frequency, corresponding to the inverse of the fundamental step size, 3.7 ± 0.6 Å.

initiation and during certain regulatory pauses^{39–42}, then the periodic release of variable amounts of scrunched DNA might also lead to discontinuous enzyme advancement. However, a more parsimonious explanation of our stepping data is that RNAP exhibits heterogeneity in individual nucleotide addition rates during transcription, causing some dwells to occur on timescales too fast to be resolved in our recordings, but maintaining tight coupling between the transcript length and position along the template.

We tested the latter explanation by performing quantitative gel

assays at $[NTP]_{eq}$ to measure transcript lengths produced by RNAP over a portion of the same DNA template used in our single-molecule studies. Analysis of these gels indicated a variability of more than an order of magnitude in local rates of NTP addition. Expected dwell-time distributions derived from gel analysis are well fitted to the actual distribution of dwell-times measured independently in single molecule records (Supplementary Fig. S3). The close correspondence between variability in next-nucleotide addition rates (measured biochemically) and variability in next-base translocation rates

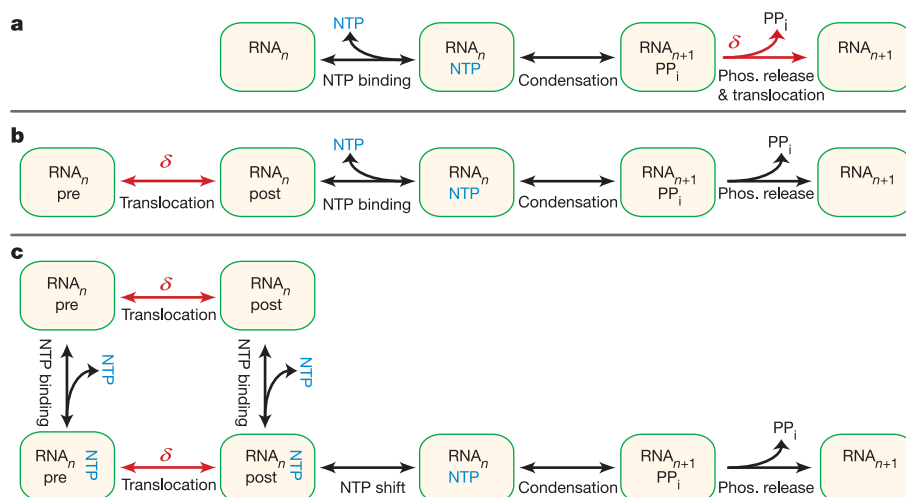


Figure 3 | Alternative kinetic models for RNAP translocation. **a**, A power stroke model where translocation (δ , red) is driven by irreversible PP_i release. **b**, A brownian ratchet model where reversible oscillation between pre- and post-translocated enzyme states can occur before NTP binding

(blue). **c**, A brownian ratchet model where translocation and NTP binding can occur in either order. This model postulates the existence of a secondary NTP site to accommodate the possibility of nucleotide binding when the enzyme is in its pre-translocated state.

(observed in single molecules) is therefore fully consistent with tight coupling between our attachment point to RNAP on the template and the length of the resulting RNA transcript. Moreover, the observation that RNAP can frequently be seen to step by single base increments (multiple examples of which are found in Fig. 2a) is incompatible with an obligate scrunching mechanism, or with any alternative mechanism where the catalytic core moves with respect to our point of enzyme attachment.

Force dependence of the nucleotide addition cycle

The application of load selectively modulates rates within the biochemical cycle involving motion. A simple Boltzmann relation describes the resulting force–velocity relationship for many mechanoenzymes^{3,43,44}:

$$v(F) = \frac{v_{\max}}{1 + \exp\left[-\frac{(F - F_{1/2})\delta}{k_B T}\right]} \quad (1)$$

where $v_{\max}$ is the velocity at large assisting load, $F_{1/2}$ is the force at which velocity reaches half its maximal value, k_B is Boltzmann's constant, T is the temperature, and δ is a parameter that represents the effective distance over which force acts. The physical interpretation of δ depends on the underlying model⁴³. In power stroke models, δ typically represents the distance from the pre-translocated position to the transition state. In such a model for RNAP inspired by ref. 29, this transition state is located between the PP_i -bound and PP_i -released states, where PP_i is pyrophosphate (Fig. 3a), and δ corresponds to some fraction of a single-base-pair separation. In brownian ratchet models, δ typically represents the characteristic

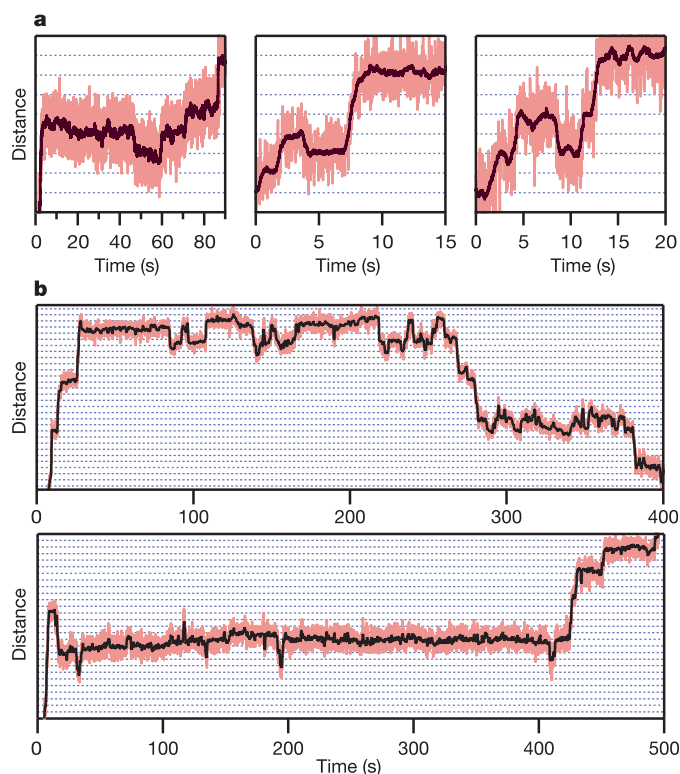


Figure 4 | RNAP backstepping and backtracking resolved at high resolution. **a**, Backstepping observed under assisting loads of 18 pN at $[\text{NTP}]_{\text{eq}}$. Molecules were occasionally found to move backward by one base pair (left and middle panels) or by two base pairs (right panel) before resuming elongation. **b**, Backtracking (>3 bp) under a hindering load of 9 pN. Molecules dwelled at specific preferred locations on the template before irreversibly backtracking (top) or recovering (bottom). Horizontal gridlines (dotted) are spaced at 3.4-Å intervals. (See Supplementary Information for a discussion.)

distance associated with fluctuations between pre- and post-translocated states. For RNAP, this distance generally corresponds to one base pair. We considered two instances of a brownian ratchet: a model inspired by ref. 32, where translocation precedes NTP binding (Fig. 3b), and a model where translocation can either precede or follow NTP binding (Fig. 3c). Because the enzyme active site is occluded by the 3' end of RNA in the pre-translocated state, the latter model requires an incoming NTP to occupy a secondary binding site before being loaded into the active site. Such a secondary binding site might represent, for example, the 'E site' observed in crystal structures of polymerase II^{45,46}, or the templated binding sites proposed in biochemical studies^{27,28,47}. These binding sites are structurally distinct, but the binding of an NTP to either type of site may be modelled by the simplified kinetic pathway shown in Fig. 3c (see also Supplementary Information).

To distinguish between the models of Fig. 3 on the basis of force–velocity behaviour, it was critical to remove load-dependent, off-pathway events from records. In particular, we observed occasional pauses associated with both backtracking¹⁹ and backstepping (Fig. 4). These off-pathway events decrease overall elongation rates by differing amounts depending on load, obscuring the on-pathway load dependence. With the improved resolution obtained, it was possible to identify and remove all rearward motions greater than 1 nm from traces. The run velocity for each molecule was computed by dividing

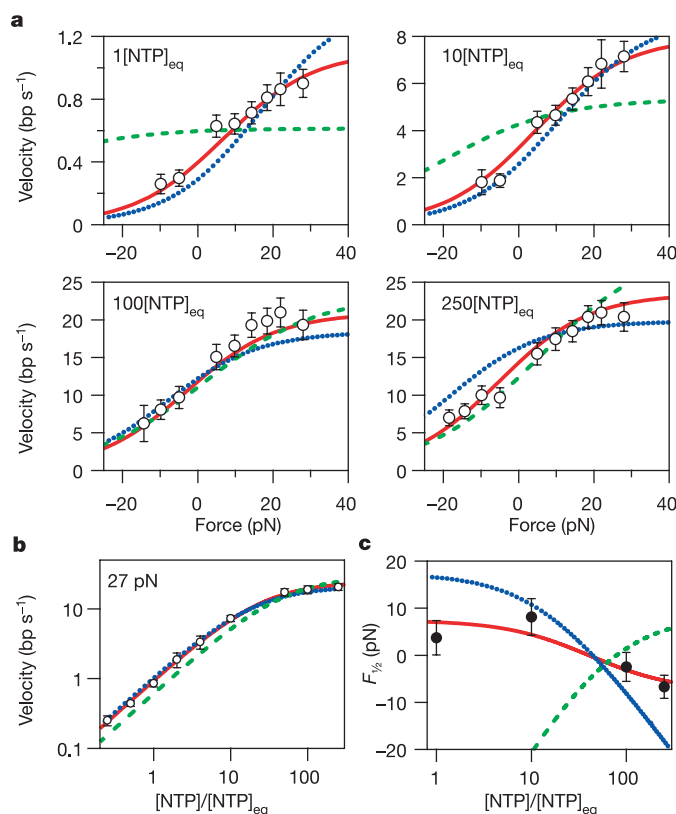


Figure 5 | Force-velocity and Michaelis-Menten relationships along with model fits. **a**, **b**, Measured force–velocity relationships for RNAP at $[\text{NTP}]_{\text{eq}}$, $10[\text{NTP}]_{\text{eq}}$, $100[\text{NTP}]_{\text{eq}}$ and $250[\text{NTP}]_{\text{eq}}$, with off-pathway events removed (see text) (**a**), and single-molecule velocity as a function of $[\text{NTP}]$ measured at 27 pN assisting load (open circles) with associated errors (see Methods) (**b**). Negative forces correspond to hindering loads; positive forces to assisting loads. Global fits to the power stroke model of Fig. 3a (green dashed line), the brownian ratchet model of Fig. 3b (blue dotted line) and the brownian ratchet model of Fig. 3c (solid red line) are shown. **c**, The $F_{1/2}$ value as a function of $[\text{NTP}]$ derived from free fits of the data in **a** to equation (1) (Supplementary Fig. S4), shown together with predictions from the three models (coloured lines) (Supplementary Fig. S5).

the total distance of advance by the total elapsed time minus any time spent in backtracked states. We note that off-pathway pauses that are insensitive to load (and therefore have no associated motion) do not affect the force dependence of nucleotide addition, because they occur with a fixed probability per unit time.

Elongation velocities were measured over a wide range of assisting and hindering loads (-18 to 28 pN) at four NTP concentrations ($1, 10, 100$ and $250 \times [\text{NTP}]_{\text{eq}}$), with an average of 32 molecules per data point (Fig. 5a). With backtracking included, the velocity interpolated to zero load at $[\text{NTP}]_{\text{eq}}$ was in excellent agreement with an independent estimate obtained through gel-based measurements. Once backtracking pauses were removed, each force–velocity curve was fit individually to equation (1). These unconstrained fits returned a characteristic distance parameter $\delta = 3.4 \pm 0.5 \text{ \AA}$ (Supplementary Fig. S4).

We generated global fits of all three models to our entire data set of velocities ($N = 40$; Fig. 5a, b). Note that more force is required to hinder elongation with increasing $[\text{NTP}]$ (that is, $F_{1/2}$ decreases; Fig. 5c), a trend opposite to that predicted for a power stroke model coupled to PP_i release. The global fit to the power stroke model generated a poor fit ($\chi^2_\nu = 6.03$; $\nu = 35$; $p(\chi^2_\nu) = 5.3 \times 10^{-27}$; five parameters; ν is the number of degrees of freedom). These findings, together with a computed distance parameter corresponding to a full base-pair displacement (rather than some fraction of a base pair expected for a power stroke model) and previous results showing that elongation velocity is insensitive to PP_i concentration⁴³, all argue against the mechanism of Fig. 3a.

A global fit of our data to the simple brownian ratchet model of Fig. 3b returned better results ($\chi^2_\nu = 2.67$; $\nu = 37$; $p(\chi^2_\nu) = 1.6 \times 10^{-7}$; three parameters), and qualitatively predicted the behaviour of $F_{1/2}$ as a function of $[\text{NTP}]$. An even better fit was obtained for the ratchet model of Fig. 3c, which invokes a secondary NTP binding site. This model predicted all unconstrained $F_{1/2}$ values to within error, and was statistically consistent with the complete data set ($\chi^2_\nu = 0.64$; $\nu = 36$; $p(\chi^2_\nu) = 0.956$; four parameters). The fit parameters suggest the presence of a small energetic penalty (~ 1 kT) associated with nucleotide binding to the secondary site when the molecule is pre-translocated, compared with the post-translocated binding energy (under standard conditions). Saturating NTP concentrations tend to ensure that the secondary binding site remains occupied and thereby bias the enzyme towards the post-translocated state. However, in contrast to the ratchet mechanism of ref. 32, hindering loads can overcome this bias by forcing the NTP-bound form into a pre-translocated state, increasing the force sensitivity at higher NTP levels. This model seems attractive for its simplicity (four free parameters), its ability to fit all available force–velocity data, and the close correspondence to recent structural and biochemical studies supplying evidence for a secondary site^{27,28,45,46}. Clearly, alternative kinetic schemes may be formulated to fit the data presented here and elsewhere.

The marked improvement in resolution obtained in this optical trapping study has led to direct measurements of base-pair stepping by an individual enzyme and supplied insights into the molecular mechanism of transcription by RNAP. Our data argue directly against any power stroke mechanism that is tightly coupled to PP_i release. Furthermore, although other recent publications have supplied independent biochemical and biophysical evidence in support of various forms of a brownian ratchet mechanism^{30,31,48}, we propose a specific model incorporating a secondary NTP binding site that is consistent with our data and others^{27,45,46}. The techniques presented here are broadly applicable. In particular, it may be possible to use our approach to relate the behaviour of a nucleic acid-based enzyme directly to the underlying DNA sequence to which it is bound, facilitating studies of sequence-dependent effects in replication, transcription and translation, and possible use in single-molecule DNA sequencing. The ability to detect motions at the ångström scale in single enzymes opens new avenues for the study of biomolecules.

METHODS

Optical trapping. The overall design of the apparatus has been described^{19,34}. Salient modifications to our apparatus include the construction of a sealed optics enclosure where helium gas at atmospheric pressure replaces ambient air and the implementation of a passive, all-optical force clamp³⁴. The force clamp was calibrated by measuring the relaxation rate of a 700-nm polystyrene bead (Bangs Labs) after release from a point ~ 300 nm from the trap centre⁴⁹. In this low-Reynolds-number regime, the velocity is proportional to the force acting on the bead: using this relationship, we found that force remained constant within 5% over a 50-nm-wide clamp region located ~ 220 nm from the trap centre. During single-molecule transcription experiments, the bead held in T_{weak} was maintained in this zero-stiffness zone by occasionally moving T_{strong} by 20 nm whenever the 600-nm bead in T_{weak} approached the outer limit of the clamp region.

For collection of force–velocity data (only), we used an active, AOD-based force clamp³³, which allowed us to alter rapidly the load on an individual enzyme during a single run, and thereby to cover the entire range of forces. Each RNAP molecule included for analysis was subjected to a cycle of hindering ($-5, -10, -14, -18$ pN) or assisting ($5, 10, 14, 18, 22, 29$ pN) loads until it either terminated or stalled. By generating data for each molecule over a range of forces, we minimized variations due to intrinsic velocity heterogeneity¹⁸.

Data analysis. Unless noted, bead displacement data were filtered with a 1-kHz, 4-pole Bessel filter, digitally acquired at 2 kHz, and median-filtered at 50 ms or 750 ms. To determine RNAP step size, we selected 37 segments of transcription through 51-Å-wide windows from 28 molecules at 18 pN or 27 pN of assisting load, then generated position histograms for each of these traces using a bin size of 0.1 Å. Histograms were autocorrelated, normalized by the number of data points, and averaged to form a global autocorrelation function. The power spectrum derived from this autocorrelation function was smoothed with a 5-point binomial filter.

To generate the graphs of Fig. 5, raw data were smoothed with an 800-ms boxcar filter and decimated to 2.5 Hz. Transcriptional pauses associated with ≥ 1 nm backward motion were removed by an automated algorithm implemented in Igor (Wavemetrics). Errors in velocity were computed as follows. First, the contribution to velocity of the positional uncertainty was estimated from the levels of statistical fluctuation occurring in regions where the enzyme had stalled. Next, the stochastic variation of RNAP motion was estimated using a randomness parameter⁵⁰ of 10 (randomness was estimated by single-molecule analysis; data not shown). Finally, a heterogeneity in the population velocity equivalent to 50% of the mean was assumed¹⁸. These three sources of error were assumed to be independent, combined in quadrature, and used to compile a weighted average of the velocity data and associated uncertainty in the mean, displayed as error bars in Fig. 5a, b.

Received 14 June; accepted 26 September 2005.

Published online 13 November 2005.

- Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **288**, 88–95 (2000).
- Purcell, T. J., Morris, C., Spudich, J. A. & Sweeney, H. L. Role of the lever arm in the processive stepping of myosin V. *Proc. Natl Acad. Sci. USA* **99**, 14159–14164 (2002).
- Altman, D., Sweeney, H. L. & Spudich, J. A. The mechanism of myosin VI translocation and its load-induced anchoring. *Cell* **116**, 737–749 (2004).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Schnitzer, M. J. & Block, S. M. Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388**, 386–390 (1997).
- Asbury, C. L., Fehr, A. N. & Block, S. M. Kinesin moves by an asymmetric hand-over-hand mechanism. *Science* **302**, 2130–2134 (2003).
- Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. Kinesin walks hand-over-hand. *Science* **303**, 676–678 (2004).
- Mallik, R., Carter, B. C., Lex, S. A., King, S. J. & Gross, S. P. Cytoplasmic dynein functions as a gear in response to load. *Nature* **427**, 649–652 (2004).
- Yasuda, R., Noji, H., Yoshida, M., Kinoshita, K. Jr & Itoh, H. Resolution of distinct rotational substeps by submillisecond kinetic analysis of F_1 -ATPase. *Nature* **410**, 898–904 (2001).
- Perkins, T. T., Dalal, R. V., Mitsis, P. G. & Block, S. M. Sequence-dependent pausing of single lambda exonuclease molecules. *Science* **301**, 1914–1918 (2003).
- van Oijen, A. M. *et al.* Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science* **301**, 1235–1238 (2003).
- Dohoney, K. M. & Gelles, J. Chi-sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules. *Nature* **409**, 370–374 (2001).
- Handa, N., Bianco, P. R., Baskin, R. J. & Kowalczykowski, S. C. Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after chi recognition. *Mol. Cell* **17**, 745–750 (2005).

14. Perkins, T. T., Li, H. W., Dalal, R. V., Gelles, J. & Block, S. M. Forward and reverse motion of single RecBCD molecules on DNA. *Biophys. J.* **86**, 1640–1648 (2004).
15. Yin, H. *et al.* Transcription against an applied force. *Science* **270**, 1653–1657 (1995).
16. Forde, N. R., Izhaky, D., Woodcock, G. R., Wuite, G. J. & Bustamante, C. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **99**, 11682–11687 (2002).
17. Adelman, K. *et al.* Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behaviour. *Proc. Natl Acad. Sci. USA* **99**, 13538–13543 (2002).
18. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**, 437–447 (2003).
19. Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426**, 684–687 (2003).
20. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
21. Li, L., Huang, H. H., Badilla, C. L. & Fernandez, J. M. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* **345**, 817–826 (2005).
22. Harada, Y. *et al.* Direct observation of DNA rotation during transcription by *Escherichia coli* RNA polymerase. *Nature* **409**, 113–115 (2001).
23. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**, 33–41 (1997).
24. Reeder, T. C. & Hawley, D. K. Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism. *Cell* **87**, 767–777 (1996).
25. Komissarova, N. & Kashlev, M. RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J. Biol. Chem.* **272**, 15329–15338 (1997).
26. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl Acad. Sci. USA* **97**, 7090–7095 (2000).
27. Zhang, C. & Burton, Z. F. Transcription factors IIF and IIS and nucleoside triphosphate substrates as dynamic probes of the human RNA polymerase II mechanism. *J. Mol. Biol.* **342**, 1085–1099 (2004).
28. Gong, X. Q., Zhang, C., Feig, M. & Burton, Z. F. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol. Cell* **18**, 461–470 (2005).
29. Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404 (2004).
30. Bai, L., Shundrovsky, A. & Wang, M. D. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J. Mol. Biol.* **344**, 335–349 (2004).
31. Bar-Nahum, G. *et al.* A ratchet mechanism of transcription elongation and its control. *Cell* **120**, 183–193 (2005).
32. Guajardo, R. & Sousa, R. A model for the mechanism of polymerase translocation. *J. Mol. Biol.* **265**, 8–19 (1997).
33. Lang, M. J., Asbury, C. L., Shaevitz, J. W. & Block, S. M. An automated two-dimensional optical force clamp for single molecule studies. *Biophys. J.* **83**, 491–501 (2002).
34. Greenleaf, W. J., Woodside, M. T., Abbondanzieri, E. A. & Block, S. M. A passive all-optical force clamp for high-resolution laser trapping. *Phys. Rev. Lett.* (in the press).
35. Rhodes, G. & Chamberlin, M. J. Ribonucleic acid chain elongation by *Escherichia coli* ribonucleic acid polymerase. I. Isolation of ternary complexes and the kinetics of elongation. *J. Biol. Chem.* **249**, 6675–6683 (1974).
36. Gelles, J., Schnapp, B. J. & Sheetz, M. P. Tracking kinesin-driven movements with nanometre-scale precision. *Nature* **331**, 450–453 (1988).
37. Block, S. M. & Svoboda, K. Analysis of high resolution recordings of motor movement. *Biophys. J.* **68**, 2305S–2395S 2395S–2415S (1995).
38. Yanagi, K., Prive, G. G. & Dickerson, R. E. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol.* **217**, 201–214 (1991).
39. Cheetham, G. M. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309 (1999).
40. Marr, M. T., Datwyler, S. A., Meares, C. F. & Roberts, J. W. Restructuring of an RNA polymerase holoenzyme elongation complex by lambdaoid phage Q proteins. *Proc. Natl Acad. Sci. USA* **98**, 8972–8978 (2001).
41. Mukherjee, S., Briebe, L. G. & Sousa, R. Discontinuous movement and conformational change during pausing and termination by T7 RNA polymerase. *EMBO J.* **22**, 6483–6493 (2003).
42. Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19**, 101–110 (2005).
43. Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
44. Block, S. M., Asbury, C. L., Shaevitz, J. W. & Lang, M. J. Probing the kinesin reaction cycle with a 2D optical force clamp. *Proc. Natl Acad. Sci. USA* **100**, 2351–2356 (2003).
45. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481–489 (2004).
46. Temiakov, D. *et al.* Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol. Cell* **19**, 655–666 (2005).
47. Holmes, S. F. & Erie, D. A. Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J. Biol. Chem.* **278**, 35597–35608 (2003).
48. Thomen, P., Lopez, P. J. & Heslot, F. Unravelling the mechanism of RNA-polymerase forward motion by using mechanical force. *Phys. Rev. Lett.* **94**, 128102 (2005).
49. Simmons, R. M., Finer, J. T., Chu, S. & Spudich, J. A. Quantitative measurements of force and displacement using an optical trap. *Biophys. J.* **70**, 1813–1822 (1996).
50. Schnitzer, M. J. & Block, S. M. Statistical kinetics of processive enzymes. *Cold Spring Harb. Symp. Quant. Biol.* **60**, 793–802 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Gelles for general discussions and continued inspiration, D. Bushnell and C. Kaplan for discussions relating to RNAP secondary binding sites, P. Fordyce, N. Guydosh, A. Meyer, A. La Porta and M. Woodside for comments on the manuscript, and R. Byer for discussions about the use of helium. W.J.G. acknowledges the support of a Predoctoral Fellowship from the NSF. This work was supported by grants to S.M.B. from the NIH-NIGMS.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.M.B. (sblock@stanford.edu).

LETTERS

Direct detection of a magnetic field in the innermost regions of an accretion disk

Jean-François Donati¹, Frédéric Paletou¹, Jérôme Bouvier² & Jonathan Ferreira²

Models^{1–5} predict that magnetic fields play a crucial role in the physics of astrophysical accretion disks and their associated winds and jets^{6,7}. For example, the rotation of the disk twists around the rotation axis the initially vertical magnetic field, which responds by slowing down the plasma in the disk and by causing it to fall towards the central star. The magnetic energy flux produced in this process points away from the disk, pushing the surface plasma outwards, leading to a wind from the disk and sometimes a collimated jet. But these predictions have hitherto not been supported by observations. Here we report the direct detection of the magnetic field in the core of the protostellar accretion disk FU Orionis⁸. The surface field reaches strengths of about 1 kG close to the centre of the disk, and it includes a significant azimuthal component, in good agreement with recent models⁵. But we find that the field is very filamentary and slows down the disk plasma much more than models predict, which may explain why FU Ori fails to collimate its wind into a jet.

Although magnetic fields have been reported in the external regions of a few protostellar disks^{9,10}, no field estimate yet exists for the innermost and densest parts of accretion disks from which most of the ejected plasma presumably originates. Obtaining constraints on the field strength and topology in accretion disk cores (such as those derived for cool star surfaces from time-resolved spectropolarimetry^{11,12}) would thus be extremely helpful for validating existing models of magnetized accretion/ejection structures around protostars and black holes. Thanks to its high mass-accretion rate^{8,13} of about 10^{-4} solar masses per year ($M_{\odot} \text{ yr}^{-1}$), FU Ori appears as a bright disk completely outshining the central protostar and is thus an obvious candidate for detecting magnetic fields in disk cores. Because models predict typical disk fields of equipartition strength (with roughly equal thermal and magnetic pressures), FU Ori should host fields of several hundred gauss in its innermost regions, within reach of modern spectropolarimetric techniques^{11,14}.

During the engineering tests of the new high-efficiency high-resolution spectropolarimeter ESPaDOnS¹⁴, we secured six observations of FU Ori at three different epochs, each consisting of one unpolarized and one circularly polarized spectrum. Using least-squares deconvolution¹¹ (LSD), we extracted the unpolarized (Stokes *I*) and circularly polarized (Stokes *V*) information from 4,700 spectral lines simultaneously, and produced mean LSD Stokes *I* and *V* profiles of FU Ori at each epoch. Whereas the unpolarized LSD profiles give information about the velocity field of the central regions of FU Ori, the circularly polarized LSD profiles give access to the putative disk magnetic field through the Zeeman signatures it generates in spectral lines. Averaging the four LSD Stokes *V* profiles recorded at the same epoch (30 November 2004), we obtain a clear Zeeman detection in the absorption lines of FU Ori (see Fig. 1); the corresponding line-of-sight magnetic field estimated from the first-order moment of the Zeeman signature¹¹ is equal to 32 ± 8 G.

This is evidence that the field we detect comes from the disk and not from either the central star or the two recently detected cool companions, each ~ 100 fainter in the V band than the accretion disk^{13,15,16}; it would otherwise require the star to exhibit a 3 kG surface-averaged line-of-sight field and thus to host a large-scale magnetic dipole of polar strength >9 kG (ref. 17), far larger than the global dipolar field component observed on both protostars¹⁸ and young low-mass stars¹². The LSD Stokes *V* profiles collected at the two other epochs (24 September 2004 and 28 November 2004) are compatible within noise level with the Zeeman signature detected on 30 November (see Fig. 1). As our observing epochs are randomly phased with respect to each other, it suggests that, at first order, the temporal fluctuations from either rotational modulation or intrinsic

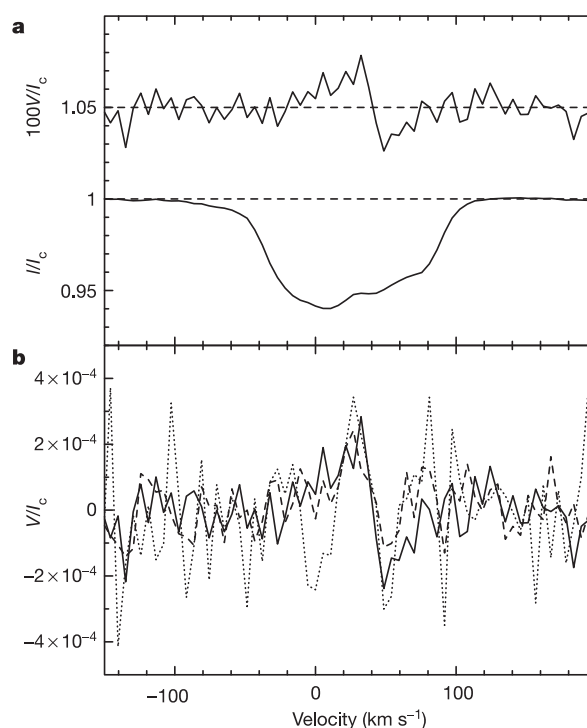


Figure 1 | Magnetic field detection in the protostellar accretion disk FU Ori. **a**, The circularly polarized (Stokes *V*) LSD profile of FU Ori (top curve, expanded by 100, normalized to the unpolarized continuum intensity I_c and shifted by +1.05) shows a clear Zeeman signature in conjunction with the unpolarized (Stokes *I*) LSD profile (bottom curve). **b**, The LSD Stokes *V* profiles obtained at the other two epochs (dashed and dotted lines) are compatible at noise level with the detected Zeeman signature (solid line), suggesting that the parent magnetic structure is grossly axisymmetric.

¹Laboratoire d'Astrophysique, Observatoire Midi-Pyrénées, F-31400 Toulouse, France. ²Laboratoire d'Astrophysique de Grenoble, Université J. Fourier, F-38041 Grenoble, France.

variability of the disk Zeeman signature are smaller than the signature itself—that is, that the parent disk magnetic structure is grossly axisymmetric.

By applying LSD with line lists corresponding to different spectral types and line strengths (see Fig. 2), we find that the derived LSD profile is closest to a conventional double-peak disk profile when using the weakest features of the G0 line list only. Although symmetric, the derived profile features a conspicuous flat-bottom shape that standard models do not predict. LSD Stokes *I* profiles produced with other line lists are all asymmetric, each at a different level; in addition to the disk profile, they include a spectral contribution strongest at spectral type A0 and blue-shifted with respect to the disk radial velocity of 27 km s^{-1} (refs 19, 20), which presumably forms in a $\sim 10,000 \text{ K}$ dense plasma ejected from the disk at speeds up to 80 km s^{-1} . This signature is probably due to the strong disk wind of FU Ori²¹. However, the detected Zeeman signature exhibits no such change with the selected line list (and in particular, no blue shift) that correlates with the spectral contribution from the wind (see Fig. 2); we thus conclude that the parent magnetic field is located within the disk only and not in the wind.

We use the LSD Stokes *I* profile produced with the partial G0 list (showing no wind contribution) and the LSD Stokes *V* signature obtained from the full G0 list (containing less noise) to investigate the velocity and magnetic fields of the disk. To this end, we constructed a simple axisymmetric disk model inspired by recent models⁵ of magnetic accretion/ejection structures. From the observation that LSD Stokes *V* profiles are two to three times narrower than their unpolarized equivalents (see Figs 1 and 3), we infer that the plasma hosting the parent magnetic field rotates at largely subkeplerian

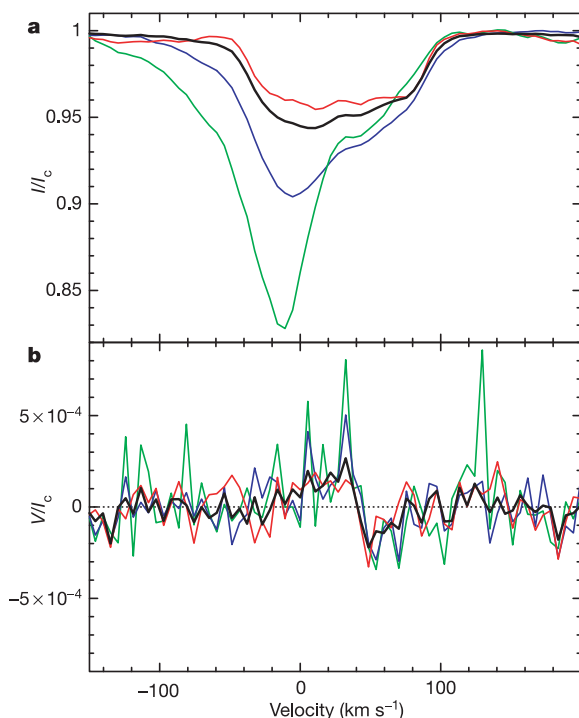


Figure 2 | Disk and wind contribution to the LSD profiles of FU Ori. **a**, Whereas the unpolarized LSD profile derived with the partial G0 line list (weak spectral features only, red line, rescaled to the size of an average line of the full G0 list) is symmetric and centred on the disk radial velocity with no additional blue-shifted absorption from the wind²¹, LSD profiles derived with the G0, A5 and A0 lists (black, blue and green lines) exhibit an increasingly larger contribution from the wind. **b**, No such changes with line list (at noise level, single pixel spikes reflecting the higher noise levels obtained with warmer line lists) are observed for the Zeeman signature, whose origin is thus attributable to the disk.

velocities. Outer disk regions rotating at keplerian velocities slow enough ($\sim 30\text{--}40 \text{ km s}^{-1}$) to produce the detected Zeeman signature are indeed too cool ($< 2,500 \text{ K}$)¹³ to contribute more than 1% to the bulk of optical lines²². Moreover, they would generate Zeeman signatures much stronger in the red (where cool regions relatively contribute more) than in the blue, in contradiction with observations (see Supplementary Fig. 1). Assuming that a small fraction of the disk plasma rotates at subkeplerian velocities also ensures that the unpolarized profile from the disk is partly filled-in and no longer features a conventional double-peak shape, but rather exhibits a flat-bottom aspect as observed.

The significant velocity shift with respect to line centre of the observed LSD Stokes *V* profile ($\sim 13 \text{ km s}^{-1}$) also indicates that the parent disk field includes a significant azimuthal component (see Fig. 3). Both vertical and azimuthal components of the field are inferred by adjusting respectively the antisymmetric and symmetric components of the detected Zeeman signature. A reasonable match to both the unpolarized profile and the circularly polarized profile from the disk is obtained when assuming that the magnetic field occupies $\sim 20\%$ of the disk plasma and that the magnetic plasma rotates two to three times slower than the local keplerian velocity; the vertical component of the derived magnetic field ($\sim 1 \text{ kG}$ at 0.05 AU) points towards the observer, while the azimuthal component (about half as strong) points in a direction opposite to the orbital rotation.

The field configuration at the surface of FU Ori resembles those predicted by theoretical models of magnetic accretion/ejection structures^{3,5,23}. We observe a steady-state magnetic structure persisting in the disk on timescales longer than the dynamical timescale.

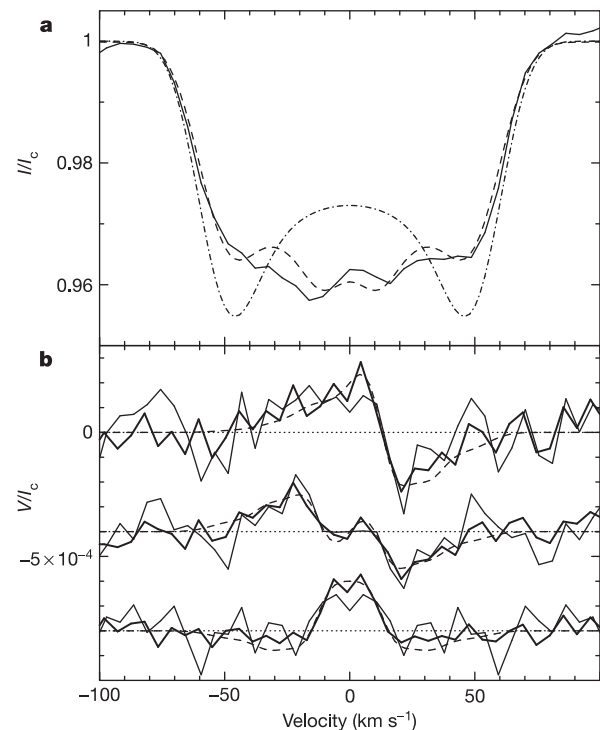


Figure 3 | Modelling the LSD profiles of FU Ori. **a**, Matching the wings of the observed unpolarized disk profile (solid line; partial G0 list) requires a disk keplerian velocity of 65 km s^{-1} at 0.05 AU (dash-dot line); matching the flat-bottom core requires that 20% of the disk plasma rotates 50–70% slower than the keplerian velocity (dashed line). **b**, The Zeeman signature (top curve) is split into its antisymmetric and symmetric components (middle and bottom curves, shifted by -4×10^{-4} and -8×10^{-4}). Fitting the observations (thick solid line, full G0 list; thin solid line, partial G0 list; dashed line, model) requires the slowly rotating disk plasma to be threaded by a 1 kG vertical field plus a 0.5 kG azimuthal field (at 0.05 AU). All profiles are shown in the disk velocity rest frame.

The detected magnetic field is of equipartition strength, and includes an azimuthal component whose direction agrees with that produced by the vertical field (presumably aligned with the large-scale field) being sheared by the disk orbital motion. The derived azimuthal to vertical field ratio at the disk surface, strongly constrained by the need to match both the symmetric and antisymmetric components of the detected Zeeman signature, is ~ 0.5 and compatible with expectations (see Fig. 7 of ref. 23), indicating that the large-scale topology is mostly dipolar rather than quadrupolar. We therefore conclude that our observations provide definite evidence, not only that accretion disks are magnetized, but also that the field configuration agrees with those devised by most recent disk/jet models. A poorer match is obtained with disk dynamo models, which predict dominantly toroidal field configurations in inner disk regions^{24,25}.

However, some differences with model predictions are observed. First, the magnetic field of FU Ori appears filamentary rather than homogeneous, and probably consists of magnetic flux tubes immersed within the bulk of the non-magnetic plasma; although plasma is at rough equipartition inside the tubes, it is essentially unmagnetized in the remaining 80% of the disk, implying an average disk magnetic energy lower than 20% of the disk thermal energy. Second, the observed magnetic plasma is slowed down much more than models expect (see Fig. 5 of ref. 23). It may indicate that the relative disk thickness h/r , with which plasma deceleration scales up (h being the disk thickness at radial distance r from disk centre), is significantly larger than the standard value of 0.1, at least in the disk core. This unexpected field property may be a hint as to why FU Ori fails to collimate its wind into a jet (as opposed to other similar objects with detected jets⁸), despite its magnetic topology being apparently favourable for jet-launching.

Our results also demonstrate that accretion disks are successful at triggering turbulent instabilities that produce enhanced radial accretion and drifts of ionized plasma through transverse field lines, as FU Ori would otherwise fail to generate a steady-state axisymmetric magnetic structure similar to what we observe. Although the nature of this instability is still enigmatic²⁶, our observations suggest that magnetic fields play a significant role in this process, making the magneto-rotational instability²⁷ and other related instability types²⁸ likely candidates.

METHODS

Spectropolarimetry with ESPaDOs. The instrument includes an achromatic polarimeter, featuring rotatable retarders and installed at the Cassegrain focus of the 3.6 m Canada-France-Hawaii Telescope atop Mauna Kea. It fibre feeds a bench-mounted high-resolution spectrograph, yielding full coverage of the 370–1,000 nm wavelength range in a single exposure¹⁴. Polarimetric exposures consist of sequences of four subexposures collected at different orientations of the polarimeter retarders, each subexposure including both orthogonal states of the selected polarization. This procedure allows the suppression of all spurious polarization signals at first order¹¹. The full efficiency of the instrument (atmosphere, telescope and detector included) is about 12%. Wavelength calibrated unpolarized and polarized spectra corresponding to each observing sequence are extracted with the dedicated software package Libre-ESpRIT, following the principles of optimal extraction²⁹.

Deriving mean line profiles with least-squares deconvolution (LSD). LSD is a cross-correlation type technique that allows us to obtain average unpolarized and polarized line profiles with largely enhanced signal to noise ratio (S/N), using simultaneously thousands of spectral features forming roughly in the same disk region¹¹. Line lists are derived from spectrum synthesis through model atmospheres³⁰. For this study, we mainly used a line list corresponding to a G0 spectral type (effective temperature of 6,000 K) and a logarithmic surface gravity of 2, which match best the spectrum of FU Ori^{19,22}. A partial G0 line list including only the weakest spectral features, as well as line lists corresponding to spectral types A5 (8,000 K) and A0 (10,000 K) are also used in the analysis. LSD profiles are found to capture well the average shape of individual lines (and in particular the conspicuous flat-bottom aspect of weak lines, see Supplementary Fig. 2). With the full G0 line list, we obtain polarized LSD profiles with noise levels of about 0.01% (relative to the unpolarized continuum level) per 5.4 km s^{-1} velocity bin, corresponding to a multiplex gain in S/N of more than 25 with respect to a single average line analysis. Higher noise levels, by factors of ~ 2 and

~ 4 respectively, are obtained with line lists corresponding to spectral types A5 and A0, containing much fewer spectral lines.

Modelling the disk profiles. The model we use is inspired by theoretical magnetic disk models⁵ and assumes axisymmetry. Ranging from 0.03 to 0.47 AU in radius, our model disk has a surface temperature (varying as $r^{-3/4}$, r being the radial distance from disk centre) of $\sim 7,000 \text{ K}$ at 0.05 AU. For a mass accretion rate^{8,13} of $10^{-4} M_{\odot} \text{ yr}^{-1}$, a turbulence parameter⁵ $\alpha_m \approx 1$ and a relative disk thickness⁵ $h/r \approx 0.1$, the number density at disk midplane (varying with $r^{-3/2}$) is $\sim 10^{17} \text{ cm}^{-3}$ at 0.05 AU. The corresponding equipartition field (which varies as $r^{-5/4}$) is 1.5 kG at 0.05 AU. The local disk brightness is assumed to scale as a blackbody. The local absorption profile is modelled as a gaussian with a full-width at half-maximum of 15 km s^{-1} and a temperature-dependent depth, whose variations are assumed gaussian with a mean and dispersion set to 6,500 and 2,500 K, respectively. We finally assume that only a fraction f of the disk plasma is magnetic, that the field in the magnetic plasma is proportional to its equipartition distribution through the disk (with scaling factors b_v and b_a for the vertical and azimuthal field components, respectively) and that the magnetic plasma rotates at a fraction k of the keplerian speed (the non-magnetic plasma being keplerian). With this simple four-parameter model, synthetic Stokes I and V profiles are obtained by integrating the Doppler-shifted spectral contributions from all disk regions (the angle of the disk rotation axis to the line-of-sight being set to 60° , as estimated from interferometry¹³).

Estimating the disk field strength and orientation. For an axisymmetric disk field configuration, the vertical and azimuthal field distributions produce Stokes V signatures that are respectively antisymmetric and symmetric with respect to the mean disk radial velocity, while the radial field distribution produces a null Zeeman signature (and thus cannot be estimated). After constructing the horizontally mirrored version V^* of the observed Zeeman signature V (with respect to the disk radial velocity of 27 km s^{-1} ; refs 19, 20), we compute the half-difference $V_a = (V - V^*)/2$ and half-sum $V_s = (V + V^*)/2$ from V and V^* and straightforwardly obtain the unique additive decomposition of $V = V_a + V_s$ into its antisymmetric and symmetric components V_a and V_s with respect to the disk radial velocity. The synthetic Zeeman signatures corresponding to the vertical and azimuthal model disk field distributions can then be matched to V_a and V_s directly (see Fig. 3).

Received 10 June; accepted 15 September 2005.

- Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
- Pringle, J. E. Accretion discs in astrophysics. *Annu. Rev. Astron. Astrophys.* **19**, 137–162 (1981).
- Blandford, R. D. & Payne, D. G. Hydromagnetic flows from accretion discs and the production of radio jets. *Mon. Not. R. Astron. Soc.* **199**, 883–903 (1982).
- Pelletier, G. & Pudritz, R. E. Hydromagnetic disk winds in young stellar objects and active galactic nuclei. *Astrophys. J.* **394**, 117–138 (1992).
- Ferreira, J. Magnetically-driven jets from Keplerian accretion discs. *Astron. Astrophys.* **319**, 340–359 (1997).
- Bridle, H. A. & Perley, A. R. Extragalactic radio jets. *Annu. Rev. Astron. Astrophys.* **22**, 319–358 (1984).
- Ray, T. P., Mundt, R., Dyson, J. E., Falle, S. & Raga, A. C. HST observations of jets from young stars. *Astrophys. J.* **468**, L103–L106 (1996).
- Hartmann, L. & Kenyon, S. J. The FU Orionis phenomenon. *Annu. Rev. Astron. Astrophys.* **34**, 207–240 (1996).
- Hutawarakorn, B. & Cohen, R. J. Magnetic structure in the bipolar outflow source G 35.2–0.74N: MERLIN spectral line results. *Mon. Not. R. Astron. Soc.* **303**, 845–854 (1999).
- Hutawarakorn, B. & Cohen, R. J. OH maser disc and magnetic field structure. *Mon. Not. R. Astron. Soc.* **357**, 338–344 (2005).
- Donati, J.-F., Semel, M., Carter, B. D., Rees, D. E. & Cameron, A. C. Spectropolarimetric observations of active stars. *Mon. Not. R. Astron. Soc.* **291**, 658–682 (1997).
- Donati, J.-F. *et al.* Dynamo processes and activity cycles of AB Dor, LQ Hya and HR 1099. *Mon. Not. R. Astron. Soc.* **345**, 1145–1186 (2003).
- Malbet, F. *et al.* New insights on the AU-scale circumstellar structure of FU Orionis. *Astron. Astrophys.* **437**, 627–636 (2005).
- Donati, J.-F. ESPaDOs: An Echelle SpectroPolarimetric Device for the Observation of Stars at CFHT. *ASP Conf. Proc.* **307**, 41–50 (2003).
- Reipurth, B. & Aspin, C. The FU Orionis binary system and the formation of close binaries. *Astrophys. J.* **608**, L68–L68 (2004).
- Wang, H., Apai, D., Henning, T. & Pascucci, I. FU Orionis: a binary star? *Astrophys. J.* **601**, L83–L86 (2004).
- Preston, G. A statistical investigation of the orientation of magnetic axes in the periodic magnetic variables. *Astrophys. J.* **150**, 547–550 (1967).
- Johns-Krull, C. M. & Gafford, A. D. New tests of magnetospheric accretion in T Tauri stars. *Astrophys. J.* **573**, 685–698 (2002).
- Herbig, G. H. Eruptive phenomena in early stellar evolution. *Astrophys. J.* **217**, 693–715 (1977).

20. Herbig, G. H., Petrov, P. P. & Duemmler, P. High-resolution spectroscopy of FU Orionis stars. *Astrophys. J.* **595**, 384–411 (2003).
21. Hartmann, L. & Calvet, N. Observational constraints on FU Ori winds. *Astron. J.* **109**, 1846–1855 (1995).
22. Kenyon, S. J., Hartmann, L. & Hewett, R. Accretion disk models for FU Orionis and V1057 Cygni: detailed comparisons between observations and theory. *Astrophys. J.* **325**, 231–251 (1988).
23. Ferreira, J. & Pelletier, G. Magnetized accretion-ejection structures III. Stellar and extragalactic jets as weakly dissipative disk outflows. *Astron. Astrophys.* **295**, 807–832 (1995).
24. Brandenburg, A., Nordlund, A., Stein, R. F. & Torkelsson, U. Dynamo-generated turbulence and large-scale magnetic fields in a Keplerian shear flow. *Astrophys. J.* **446**, 741–754 (1995).
25. von Rekowski, B., Brandenburg, A., Dobler, W. & Shukurov, A. Structured outflow from a dynamo active accretion disc. *Astron. Astrophys.* **398**, 825–844 (2003).
26. Dubrulle, B. *et al.* An hydrodynamic shear instability in stratified disks. *Astron. Astrophys.* **429**, 1–13 (2005).
27. Balbus, S. A. & Hawley, J. F. A powerful local shear instability in weakly magnetised discs. *Astrophys. J.* **376**, 214–233 (1991).
28. Keppens, R., Casse, F. & Goedbloed, J. P. Waves and instabilities in accretion disks: magnetohydrodynamic spectroscopic analysis. *Astrophys. J.* **569**, L121–L126 (2002).
29. Horne, K. An optimal extraction algorithm for CCD spectroscopy. *Publ. Astron. Soc. Pacif.* **98**, 609–617 (1986).
30. Kurucz, R. L. *Atlas 9 Stellar Atmospheres and Programs* [CD set] (Smithsonian Astrophysical Observatory, Cambridge, USA, 1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the CFHT staff for help during the engineering runs of ESPaDOs. We are grateful to F. Rincon and A. C. Cameron for comments on earlier versions of the manuscript. This Letter is based on observations obtained at the Canada-France-Hawaii Telescope (CFHT), which is operated by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.-F.D. (donati@ast.obs-mip.fr).

LETTERS

Experimental implementation of heat-bath algorithmic cooling using solid-state nuclear magnetic resonance

J. Baugh¹, O. Moussa¹, C. A. Ryan¹, A. Nayak^{1,2,3} & R. Laflamme^{1,3}

The counter-intuitive properties of quantum mechanics have the potential to revolutionize information processing by enabling the development of efficient algorithms with no known classical counterparts^{1,2}. Harnessing this power requires the development of a set of building blocks³, one of which is a method to initialize the set of quantum bits (qubits) to a known state. Additionally, fresh ancillary qubits must be available during the course of computation to achieve fault tolerance^{4–7}. In any physical system used to implement quantum computation, one must therefore be able to selectively and dynamically remove entropy from the part of the system that is to be mapped to qubits. One such method is an ‘open-system’ cooling protocol in which a subset of qubits can be brought into contact with an external system of large heat capacity. Theoretical efforts^{8–10} have led to an implementation-independent cooling procedure, namely heat-bath algorithmic cooling. These efforts have culminated with the proposal of an optimal algorithm, the partner-pairing algorithm, which was used to compute the physical limits of heat-bath algorithmic cooling¹¹. Here we report the experimental realization of multi-step cooling of a quantum system via heat-bath algorithmic cooling. The experiment was carried out using nuclear magnetic resonance of a solid-state ensemble three-qubit system. We demonstrate the repeated repolarization of a particular qubit to an effective spin-bath temperature, and alternating logical operations within the three-qubit subspace to ultimately cool a second qubit below this temperature. Demonstration of the control necessary for these operations represents an important step forward in the manipulation of solid-state nuclear magnetic resonance qubits.

Nuclear magnetic resonance (NMR)-based ensemble quantum information processing (QIP) devices have provided excellent test beds for controlling non-trivial numbers of qubits^{12–15}. A solid-state NMR QIP architecture builds on this success by incorporating the essential features of the liquid-state devices while offering the potential to reach unit polarization and thus control more qubits^{15,16}. In this architecture, the abundant nuclear spins with polarization P form a large-heat-capacity spin-bath that can be either coupled to, or decoupled from, a dilute, embedded ensemble of spin-labelled isotopomers that comprise the qubit register. Bulk spin-cooling procedures such as dynamic nuclear polarization are well known and capable of reaching polarizations near unity^{15,17}. This architecture is one realization within a large class of possible solid-state QIP systems in which coherently controlled qubits can be brought into contact with an external system that behaves as a heat bath. The principles and methods applied in solid-state NMR QIP will therefore apply to many other systems. An additional motivation is development of control techniques that future quantum devices

will harness. For this experiment, we develop a novel technique to implement the controlled qubit–bath interaction, and also report the first application of strongly modulating pulses¹⁸ to solid-state NMR for high-fidelity, coherent qubit control.

The three-qubit quantum information processor used here is formed by the three spin-1/2 ¹³C nuclei of isotopically labelled malonic acid molecules, occupying a dilute fraction of lattice sites in an otherwise unlabelled single crystal of malonic acid (unlabelled, with the exception of naturally occurring ¹³C isotopes at the rate of 1.1%). The concentration of labelled molecules was 3.2%. Malonic acid also contains abundant spin-1/2 ¹H nuclei, which comprise the heat-bath. Figure 1 shows the ¹H-decoupled, ¹³C-NMR spectrum for the crystal (and crystal orientation) used in this work. The spectrum shows the NMR absorption peaks of both the qubit spins (quartets) and natural abundance ¹³C spins (singlets), the latter being inconsequential for QIP purposes. The table in Fig. 1 lists the parameters of the ensemble qubit hamiltonian obtained from fitting the spectrum, and also includes couplings involving the methylene protons calculated for this crystal orientation from the known crystal structure¹⁹. Experiments were performed at room temperature at a static magnetic field strength of 7.1 T, where the thermal ¹H polarization is $P_H \approx 2.4 \times 10^{-5}$.

In this orientation, the methylene carbon C_m has a dipolar coupling of 19 kHz to H_{m1} of the methylene ¹H pair, whereas no other ¹³C–¹H dipolar coupling in the system is larger than 2 kHz (see Fig. 1 for atom nomenclature). Therefore, a spin-exchange hamiltonian of the form

$$H_{\text{ex}} = \sum_{j \in C, k \in H} \frac{D_{jk}}{3} \frac{\sigma_z^j \sigma_z^k + \sigma_y^j \sigma_y^k + \sigma_x^j \sigma_x^k}{2} \quad (1)$$

that couples the two nuclear species will generate dynamics dominated by the large C_m – H_{m1} coupling at short times (the D_{jk} are ¹³C–¹H dipolar couplings, indices j, k run over ¹³C, ¹H nuclei, respectively, and σ_β^α is the β -axis Pauli operator for spin α). Starting from the natural coupling hamiltonian, $H_{\text{nat}} = \sum_{j \in C, k \in H} D_{jk} \sigma_z^j \sigma_z^k / 2$, we

applied a multiple-pulse ‘time-suspension’ sequence²⁰ synchronously to both ¹³C and ¹H spins to create the effective spin-exchange hamiltonian (in the toggling frame), to lowest order in the Magnus expansion of the average hamiltonian²¹. Application of the sequence for the C_m – H_{m1} exchange period $\tau = 3/(4 \times 19 \text{ kHz}) \approx 40 \mu\text{s}$ results in an approximate swap gate (state exchange) between the C_m and H_{m1} spins. With an initial bulk ¹H polarization P_H , this procedure yields a selective dynamic transfer of polarization $P' = \eta P_H$ to C_m , where $0 \leq |\eta| \leq 1$ and ideally $|\eta| = 1$. We define the effective spin-bath temperature to be that which corresponds to the experimentally

¹Institute for Quantum Computing, ²Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. ³Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada.

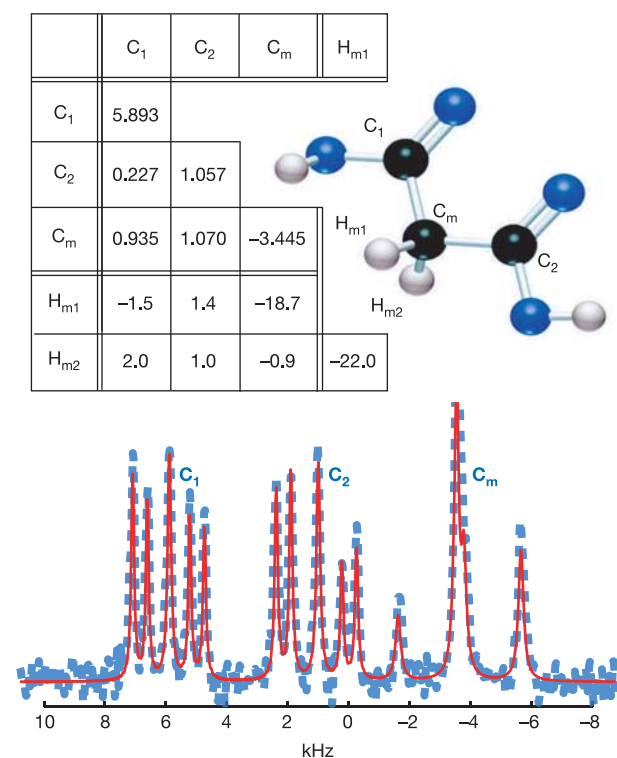


Figure 1 | Characteristics of the dilute $3\text{-}^{13}\text{C}$ malonic acid spin system. Bottom, ^1H -decoupled, ^{13}C spectrum near the [010] orientation with respect to the static magnetic field. The blue-dashed line is the experimental NMR absorption spectrum, and the solid red line is a fit. Multiplet assignments are indicated by the labels C_1 , C_2 and C_m . The central peaks in each multiplet correspond to natural abundance ^{13}C in the sample, which are inconsequential for QIP purposes. The peak height differences in the $3\text{-}^{13}\text{C}$ molecule peaks indicate the strong coupling regime, that is, the $^{13}\text{C}\text{-}^{13}\text{C}$ intramolecular dipolar couplings are significant compared to the relative chemical shifts. Top, table showing the ^{13}C rotating-frame hamiltonian parameters (chemical shifts along diagonal; dipolar coupling strengths off-diagonal; all values in kHz) obtained from the spectral fit. It also includes calculated dipolar couplings involving the methylene protons based on the atomic coordinates¹⁹ and the crystal orientation obtained from the spectral fit.

obtained P' under this procedure, and refer to this transfer as a refresh operation. We obtained $P' \approx 0.83P_H$ experimentally, and found that repeated refresh operations showed no loss in efficiency given at least a 6 ms delay for $^1\text{H}\text{-}^1\text{H}$ equilibration. However, we observed a decay of P_H as a function of the number of repetitions, due to accumulated control errors, which lead to an identical loss in the refresh polarization.

The experiment consists of the first six operations of the partner-pairing algorithm (PPA) on three qubits: three refresh operations, and three permutation gates that operate on the qubit register. This is described in the quantum circuit diagram of Fig. 2. During the register operations, the ^1H polarization is first rotated into the transverse plane, and then 'spin-locked' by a strong, phase-matched radio frequency (r.f.) field that both preserves the bulk ^1H polarization and decouples the $^1\text{H}\text{-}^{13}\text{C}$ dipolar interactions. As $^1\text{H}\text{-}^1\text{H}$ dipolar interactions are merely scaled by a factor $-1/2$ under spin-locking, H_{m1} is allowed to equilibrate with the bulk ^1H nuclei via spin diffusion. This occurs on a timescale longer than the transverse dephasing time ($T_2(H_m) \approx 100 \mu\text{s}$), but much shorter than the spin-lattice relaxation time ($T_1^H \approx 50 \text{ s}$) of H_{m1} . Hence, H_{m1} plays the role of the fast-relaxing qubit described in the protocol of ref. 11. The first two register operations are swap gates; the third is a three-bit compression (3BC) gate⁸⁻¹⁰ that boosts the polarization of the first qubit, C_1 , at the expense of the polarizations of the other two qubits.

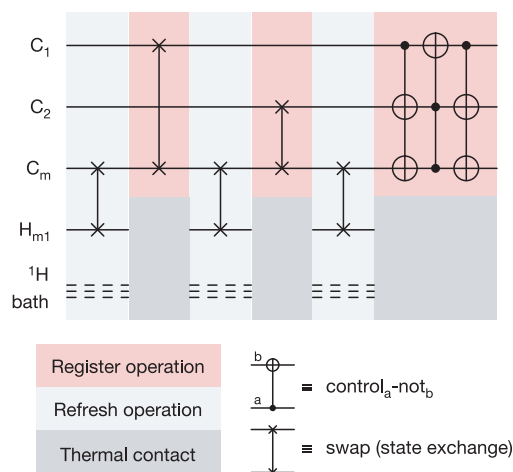


Figure 2 | Schematic quantum circuit diagram of the implemented protocol. Time flows from left to right. The three-bit compression (3BC) gate is shown here decomposed as control-not gates and a control-control-not (Toffoli) gate. The gate sequence corresponds to the first six steps of the partner-pairing algorithm¹¹ on three qubits. The input state is a collective polarization P_H of the bulk ^1H . The refresh operation is approximately $40 \mu\text{s}$ in duration, whereas the register operations are between 0.7 and 1.3 ms in duration. Thermal contact takes place during ^1H spin-locking pulses that begin just before the register operations, and extend an additional 12 ms after each operation. H_{m1} can be thought of as an additional 'special purpose' qubit in this experiment; despite non-selective ^1H control (due to bulk hydrogenation), the refresh and thermal contact operations could be performed using collective ^1H control. Thus, H_{m1} serves as a fast-relaxing 'qubit' and the bulk ^1H -bath as a thermal bath of large heat capacity.

Ideally, the protocol builds a uniform polarization on all three qubits corresponding to the bath polarization (first five steps), then selectively transfers as much entropy as possible from the first qubit to the other two (last step). The last step (3BC) leads to a polarization boost by a factor of $3/2$ on the first qubit. Subsequently, the heated qubits can be re-cooled to the spin-bath temperature, and the compression step repeated, iteratively, until the asymptotic value of the first-bit polarization is reached. This limiting polarization depends only on the number of qubits and the bath polarization¹¹, and is ideally $P(C_1) = 2P'$ for three qubits (for n qubits it is $2^{n-2}P'$ in the regime $P' \ll 2^{-n}$, and 1.0 in the regime $P' \gg 2^{-n}$ (refs 11, 22)). The first six steps carried out here should yield a polarization of $1.5P'$ on C_1 , assuming ideal operations.

The control operations performed here are quantum control operations: state-independent unitary rotations in the Hilbert space. However, it should be noted that the heat-bath algorithmic cooling gates are all permutations that map computational basis states to other computational basis states. Therefore, gate fidelities were measured with respect to correlation with these known states, rather than the manifold of generic quantum states. We took advantage of this property to further optimize the control parameters of the ^{13}C gates (register operations) for the state-specific transformations of the protocol. These operations were carried out using numerically optimized control sequences referred to as strongly modulating pulses¹⁸. Such pulses drive the system strongly at all times, such that the average r.f. amplitude is comparable to, or greater than, the magnitude of the internal hamiltonian. This allows inhomogeneities in the ensemble qubit hamiltonian to be efficiently refocused, so that ensemble coherence is better maintained throughout the gate operations.

In this set of experiments, the ^{13}C qubit spins are initialized to infinite temperature (a preceding broadband ^{13}C $\pi/2$ excitation pulse is followed by a dephasing period in which ^1H dipolar fields effectively dephase the ^{13}C polarization). Following the fifth step,

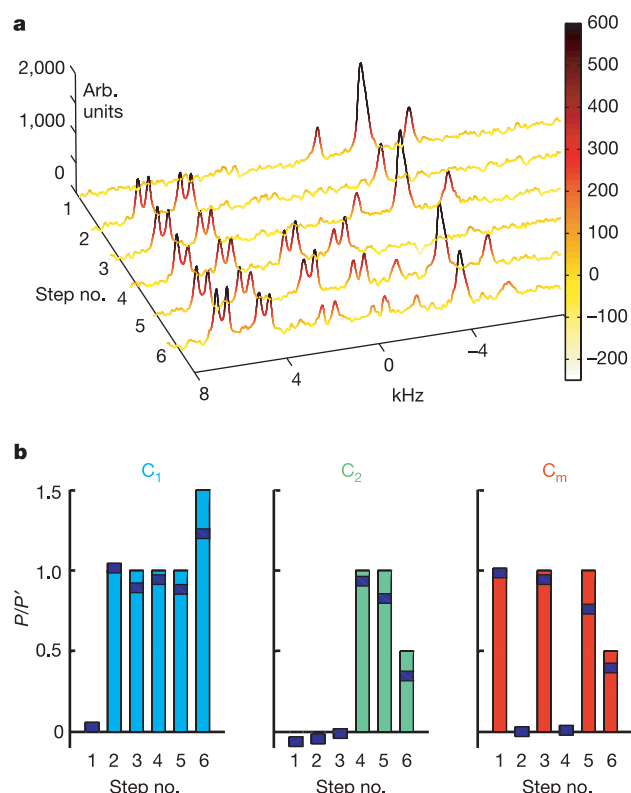


Figure 3 | Experimental results in terms of ^{13}C spectra and their integrated intensities. **a**, Readout spectra obtained following each of the six steps in the protocol. A colour scale indicates peak intensities, which are in arbitrary units. The integrated peak intensities for each multiplet correspond to the ensemble spin polarizations. The natural abundance C_m signal that appears at each refresh step (adding to the intensity of the central peaks) should be ignored; we are only interested in the part of the signal arising from the $3\text{-}^{13}\text{C}$ qubit molecules, which can be seen clearly in the C_1 and C_2 spectral regions. **b**, Bars indicate ideal qubit polarizations at each step; experimental values obtained from integration of the above spectra are shown as shaded bands, whose thickness indicates experimental uncertainty.

polarizations (in units of P') of 0.88, 0.83 and $0.76 (\pm 0.03)$ are built up on C_1 , C_2 and C_m , respectively. The final 3BC operation yields $P(C_1)/P' = 1.22 \pm 0.03$, an increase of 48% compared to the average polarization (0.82) following step five. Despite control imperfections that effectively heat the qubits at each step, we are able to cool the C_1 qubit ensemble well below the effective ^1H spin-bath temperature.

The results are summarized in Fig. 3; in Fig. 3a are shown the spectral intensities corresponding to ^{13}C spin polarizations following each of the six steps, and in Fig. 3b the integrated intensities are graphed in comparison with the ideal values. We note that the overall fidelity of the experiment, $F = 1.22/1.50 = 0.81$, implies an error per step of 3.7%. This error rate is only about a factor of two larger than the average error per two-qubit gate obtained in a benchmark liquid-state NMR QIP experiment¹². Furthermore, the state-correlation fidelity of the 3BC gate over the polarizations on all three qubits is 0.96 ± 0.03 . From Fig. 3b, it can be seen that the fidelity of the refresh operation drops off roughly quadratically in the number of steps; this is consistent with the loss of bulk ^1H polarization due to pulse imperfections both in the multiple-pulse refresh operations and in the spin-locking sequence. As the broadband pulses have been optimized for flip-angle in these sequences, we suspect that the remaining errors are mainly due to switching transients that occur in the tuned r.f. circuitry of the NMR probe head, and to a lesser extent off-resonance and finite pulse-width effects that modify the average hamiltonian²⁰. Similar effects lead to imperfect fidelity of the ^{13}C control. With suitable improvements to the resonant circuit response

and by incorporating numerical optimization of the multiple-pulse refresh operations, we expect that several iterations of the protocol could be carried out and that the limiting polarization of $2P'$ could be approached in this system. The same methodologies should also be applicable in larger qubit systems with similar architecture. For a six-qubit system using the PPA, a bath polarization $P' > 0.2$ would be sufficient, in principle, to reach a pure state on one qubit²². Such bulk nuclear polarizations are well within reach via well-known dynamic nuclear polarization techniques¹⁷; for example, unpaired electron spins at defects (g -factor = 2) in a field of 3.4 T and at temperature 4.2 K are polarized to 0.5.

This work demonstrates that solid-state NMR QIP devices could be used to implement active error correction. Given a bath polarization near unity, the refresh operation implemented here would constitute the dynamic resetting of a chosen qubit. This would allow a new NMR-based test bed for the ideas of quantum error correction and for controlled open-system quantum dynamics in the regime of high state purity and up to approximately 20 qubits.

METHODS

NMR experiments were carried out at room temperature on a Bruker Avance solid-state spectrometer operating at a field of 7.1 T, and a purpose-built dual channel r.f. probe. The sample coil had an inner diameter of 3 mm, and the employed $\pi/2$ broadband pulse lengths were 1.25 μs and 0.75 μs for ^{13}C and ^1H , respectively. The sample was a $4 \times 1.5 \times 1.5 \text{ mm}^3$ single crystal of malonic acid grown from aqueous solution with a 3.2% molecular fraction of $3\text{-}^{13}\text{C}$ labelled molecules. Spectra were obtained by signal averaging for 80 scans. The proton spin-lattice relaxation time was $T_1^H = 50 \text{ s}$, so the delay between scans was set to $6T_1^H = 300 \text{ s}$. The average ^{13}C free-induction dephasing rate (with ^1H decoupling) was $1/T_2^* = (2 \text{ ms})^{-1}$ in our sample. This rate is dominated by the ensemble dispersion of chemical shifts, much of which is effectively removed by the control operations. Design of the strongly modulating ^{13}C pulses followed very closely the methodology described in ref. 18, and penalty functions were adjusted to favour average r.f. amplitudes comparable to or greater than the magnitude of the ^{13}C rotating-frame hamiltonian. These pulses were optimized and simulated over a five-point distribution of r.f. amplitude corresponding to the measured distribution over the spin ensemble ($\sigma = 6.2\%$ in r.f. amplitude). The 'time-suspension' sequence applied synchronously to ^{13}C and ^1H was a 12-pulse subsequence of the Cory 48-pulse sequence²⁰. The delays between pulses were adjusted so that the total length of the sequence was 40 μs . ^1H spin-locking/decoupling was carried out at an r.f. amplitude of 250 kHz. The spectra in Fig. 3 were obtained by applying a $\pi/2$ broadband pulse to read out the spin polarizations. The absolute value of the refresh polarization P' was determined by comparing the initial refresh polarization on C_m and the thermal equilibrium ^{13}C polarization P_C measured in a separate experiment. These yield the ratio of P' to P_C , and P' to P_H using the fact that $P_H = 3.98P_C$.

Received 14 July; accepted 22 September 2005.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
- Everitt, H., Havel, T. F. & Cory, D. G. (eds) Experimental aspects of quantum computing. *Quant. Inform. Process.* 3(1–5; special issue), 1–308 (2004).
- DiVincenzo, D. P. The physical implementation of quantum computation. *Fortschr. Phys.* 48, 771–793 (2000).
- Knill, E., Laflamme, R. & Zurek, W. H. Resilient quantum computation. *Science* 279, 342–345 (1998).
- Kitaev, A. Y. in *Quantum Communication and Computing and Measurement* (eds Holevo, A. S., Hirota, O. & Caves, C. M.) 181–188 (Plenum, New York, 1997).
- Aharonov, D. & Ben-Or, M. in *Proc. 29th Annual ACM Symp. on Theory of Computing* (eds Leighton, F. T. & Shor, P.) 176–188 (ACM Press, New York, 1997).
- Preskill, J. Reliable quantum computers. *Proc. R. Soc. Lond. A* 454, 385–410 (1998).
- Schulman, L. J. & Vazirani, U. V. in *Proc. 31st Annual ACM Symp. on Theory of Computing* (eds Vitter, J. S., Larmore, L. & Leighton, T.) 322–329 (ACM Press, New York, 1999).
- Boykin, P. O., Mor, T., Roychowdhury, V., Vatan, F. & Vrijen, R. Algorithmic cooling and scalable NMR quantum computers. *Proc. Natl Acad. Sci. USA* 99, 3388–3393 (2002).
- Fernandez, J. M., Lloyd, S., Mor, T. & Roychowdhury, V. Algorithmic cooling of spins: a practicable method for increasing polarization. *Int. J. Quant. Inform.* 2, 461–467 (2004).

11. Schulman, L. J., Mor, T. & Weinstein, Y. Physical limits of heat-bath algorithmic cooling. *Phys. Rev. Lett.* **94**, 120501 (2005).
12. Knill, E., Laflamme, R., Martinez, R. & Tseng, C.-H. An algorithmic benchmark for quantum information processing. *Nature* **404**, 368–370 (2000).
13. Gershenfeld, N. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).
14. Cory, D. G., Price, M. D. & Havel, T. F. Nuclear magnetic resonance spectroscopy: An experimentally accessible paradigm for quantum computing. *Physica D* **120**, 82–101 (1998).
15. Cory, D. G. *et al.* NMR based quantum information processing: achievements and prospects. *Fortschr. Phys.* **48**, 875–907 (2000).
16. Leskowitz, G. M., Ghaderi, N., Olsen, R. A. & Mueller, L. J. Three-qubit nuclear magnetic resonance quantum information processing with a single-crystal solid. *J. Chem. Phys.* **119**, 1643–1649 (2003).
17. Abragam, A. & Goldman, M. *Nuclear Magnetism: Order and Disorder* 339–393 (Clarendon, Oxford, 1982).
18. Fortunato, E. M. *et al.* Design of strongly modulating pulses to implement precise effective Hamiltonians for quantum information processing. *J. Chem. Phys.* **116**, 7599–7606 (2002).
19. Jagannathan, N. R., Rajan, S. S. & Subramanian, E. Refinement of the crystal structure of malonic acid, $C_3H_4O_4$. *J. Chem. Crystallogr.* **24**, 75–78 (1994).
20. Cory, D. G., Miller, J. B. & Garroway, A. N. Time-suspension multiple-pulse sequences: applications to solid-state imaging. *J. Magn. Res.* **90**, 205–213 (1990).
21. Haeberlen, U. *High Resolution NMR in Solids: Selective Averaging* Suppl. 1, *Advances in Magnetic Resonance* 64–69 (Academic, New York, 1976).
22. Moussa, O. *On Heat-bath Algorithmic Cooling and its Implementation in Solid-state NMR*. Page 20, M.Sc. thesis, Univ. Waterloo (2005); available at <http://www.iqc.ca/~omoussa/work/thesis/>.

Acknowledgements We thank D. G. Cory, T. F. Havel and C. Ramanathan for discussions and use of NMR simulation code; W. P. Power, M. J. Ditty and N. J. Taylor for facility use and experimental assistance; CIAR, ARDA and NSERC for support. O.M. acknowledges the Ontario Ministry of Training, Colleges and Universities for support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.B. (baugh@iqc.ca).

LETTERS

Nodal quasiparticle in pseudogapped colossal magnetoresistive manganites

N. Mannella^{1,2}, W. L. Yang^{1,2}, X. J. Zhou^{1,2}, H. Zheng³, J. F. Mitchell³, J. Zaanen^{1,4}, T. P. Devereaux⁵, N. Nagaosa^{6,7}, Z. Hussain² & Z.-X. Shen¹

A characteristic feature of the copper oxide high-temperature superconductors is the dichotomy between the electronic excitations along the nodal (diagonal) and antinodal (parallel to the Cu–O bonds) directions in momentum space, generally assumed to be linked to the ‘d-wave’ symmetry of the superconducting state. Angle-resolved photoemission measurements in the superconducting state have revealed a quasiparticle spectrum with a d-wave gap structure that exhibits a maximum along the antinodal direction and vanishes along the nodal direction¹. Subsequent measurements have shown that, at low doping levels, this gap structure persists even in the high-temperature metallic state, although the nodal points of the superconducting state spread out in finite ‘Fermi arcs’². This is the so-called pseudogap phase, and it has been assumed that it is closely linked to the superconducting state, either by assigning it to fluctuating superconductivity³ or by invoking orders which are natural competitors of d-wave superconductors^{4,5}. Here we report experimental evidence that a very similar pseudogap state with a nodal–antinodal dichotomous character exists in a system that is markedly different from a superconductor: the ferromagnetic metallic groundstate of the colossal magnetoresistive bilayer manganite $\text{La}_{1.2}\text{Sr}_{1.8}\text{Mn}_2\text{O}_7$. Our findings therefore cast doubt on the assumption that the pseudogap state in the copper oxides and the nodal–antinodal dichotomy are hallmarks of the superconductivity state.

$\text{La}_{1.2}\text{Sr}_{1.8}\text{Mn}_2\text{O}_7$ (LSMO) is a prototypical bilayer manganite that exhibits the colossal magnetoresistance (CMR) effect—the extremely large drop in resistivity induced by application of a magnetic field near the Curie temperature (T_C)^{6,7}. The CMR effect exploits a metal–insulator transition between a low-temperature ferromagnetic-metallic ground state and a high-temperature paramagnetic-insulating phase. The nature of the ferromagnetic-metallic ground state in LSMO remains highly controversial. On the one hand, the underlying Fermi surface and band structure have clear resemblance to those of the copper oxides⁸ (Lin, H., Sahrakorpi, S., Barbiellini, B. & Bansil, A., personal communication on full potential band structure and Fermi surface computations for $x = 0.4$). On the other hand, previous angle-resolved photoemission (ARPES) investigations revealed no quasiparticle peak, a suppression of spectral weight at the Fermi level (E_F) (the pseudogap), and an unusually light effective mass, a factor of two lighter than the calculated band structure value. This last observation is puzzling given the general expectation of strong interactions in the manganites. Moreover, the value for the in-plane conductivity calculated with the ARPES parameters is nearly one order of magnitude higher than that measured by transport^{9–11}.

Our ARPES experiments resolve the controversy of the low-temperature ferromagnetic-metallic groundstate of LSMO by

demonstrating that its electronic structure is strikingly similar to that found in the pseudogap phase of the copper oxide high-temperature superconductors (HTSC). At 20 K, well below $T_C \approx 120$ K, our data show a small but well-defined quasiparticle

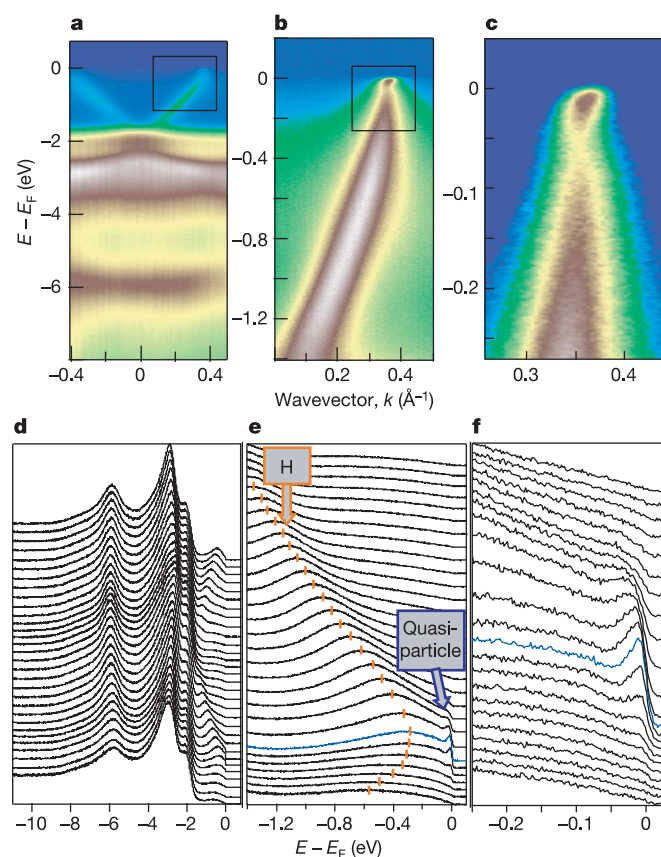


Figure 1 | Data collected along the (0,0)–(π , π) nodal direction at various magnifications. **a**, **d**, Image plot and EDC of the valence band around the Γ point. The Mn e_g states show a ~ 1.8 eV dispersion, ~ 0.4 eV wider than that predicted by LDA calculations⁸ (Lin, H., Sahrakorpi, S., Barbiellini, B. & Bansil, A., personal communication). **b**, Magnification of the box in **a** and stacks of EDC (in **e**) of the e_g band, showing the break-up of the spectral weight into a small quasiparticle peak which crosses E_F at a Fermi wavevector $k_F \approx 0.37 \text{ \AA}^{-1}$, and a rapidly dispersing broad peak (the hump, H). **c**, Magnification of the box in **b** and relative EDC (**f**). The data have been collected at 20 K with photon energy $h\nu = 42$ eV and the light polarization vector lying in the sample plane.

¹Departments of Physics, Applied Physics, and Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford, California 94305, USA. ²Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ³Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁴Instituut Lorentz for Theoretical Physics, Leiden University, POB 9506, 2300 RA Leiden, The Netherlands. ⁵Department of Physics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. ⁶CREST, Department of Applied Physics, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ⁷Correlated Electron Research Center, AIST, Tsukuba 305-8562, Japan.

peak near the (0,0) to (π,π) nodal direction (Fig. 1), with dispersion yielding an effective mass almost six times heavier than the band structure value (Fig. 2). The weight of the quasiparticle peak diminishes rapidly away from the nodal direction while crossing over to well-nested ghost-like Fermi-surface segments (Fig. 4). This nodal–antinodal dichotomy, in terms of anisotropy, spectral line-shapes and the nested ‘ghost-like’ Fermi surface away from the node, is almost identical at a phenomenological level to the one found in the underdoped $\text{Ca}_{2-x}\text{Na}_x\text{CuO}_2\text{Cl}_2$ HTSC¹².

Figure 1 shows the data collected along the nodal direction at various magnifications. Because we are not dealing with a d-wave superconductor, the term ‘nodal direction’ does not have any special meaning in the context of manganites. Nevertheless, we use this terminology because it is useful for comparisons with the copper oxides. A striking feature is the particularly ‘clean’ band seen to terminate at E_F . This corresponds to the Mn e_g states, with a bandwidth of ~ 1.8 eV and ~ 0.4 eV wider than the band structure prediction using a local density approximation (LDA)⁸ (Lin, H., Sahrakorpi, S., Barbiellini, B. & Bansil, A., personal communication). To our knowledge, such a well-defined and wide dispersion has not previously been observed in any transition metal oxide. The sharp energy scale near 300–400 meV, ubiquitous in HTSC owing to antiferromagnetic interactions, is absent here^{13,14}.

Upon magnification, the image plots (Fig. 1b, c) and the ‘peak–dip–hump’ structure in the energy distribution curves (EDC) (Fig. 1e, f) show some of the canonical signatures of a strong coupling to bosonic modes close to the polaronic limit, albeit highly momentum-dependent^{15–18}. The spectral weight is split into two branches: (1) a small but well-defined quasiparticle peak with a very narrow dispersion below ~ 50 meV and (2) a broad ‘hump’ peak whose maximum disperses along a roughly parabolic band from ~ 300 meV below E_F down to the Γ point at ~ 1.8 eV.

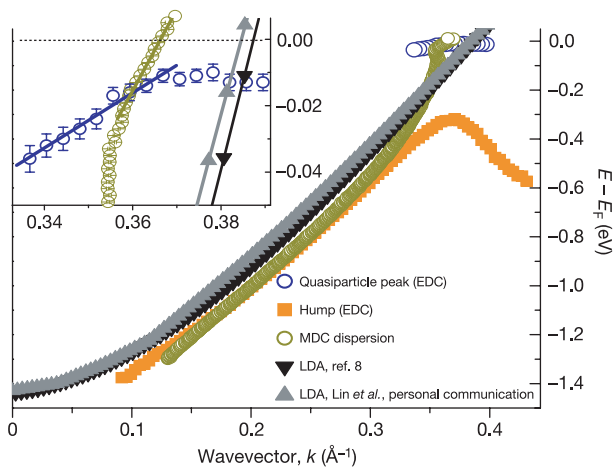


Figure 2 | Dispersion of the e_g band. Aside from a rigid shift, there seems to be a good correspondence for energies higher than ~ 300 – 400 meV between the e_g band dispersions calculated with LDA⁸ (Lin, H., Sahrakorpi, S., Barbiellini, B. & Bansil, A., personal communication) and extracted with the MDC analysis of the data shown in Fig. 1b. Also shown are the positions of the broad peak (hump) and the quasiparticle peak determined from the EDC shown in Figs 1e and f. The inset shows the magnification of the region close to the Fermi level crossing. The solid lines indicate the results of the fittings. If not shown, the error bars (± 1 s.d.) are smaller than the symbol's size. The MDC analysis for energies approaching E_F leads to an S-shaped dispersion with two kinks near 50 and 100 meV. This S-shaped dispersion is an artefact resulting from the way the MDC analysis handles the Englesberg–Schrieffer two-poles solution¹⁹: namely the distribution of spectral weight for energies ranging from ~ 50 meV to the maximum of the broad hump peak. Nevertheless, the energy scale revealed by the 50 meV kink is still meaningful, as it marks the asymptotic energy separating the coherent quasiparticle peak from the hump, as also shown in Fig. 1c.

In Fig. 2 we compare the LDA results in ref. 8 and in the work of H. Lin, S. Sahrakorpi, B. Barbiellini & A. Bansil (personal communication) with the e_g band dispersion obtained from the EDC and momentum distribution curves (MDC) analyses of the data shown in Fig. 1. The high-energy incoherent branch (the hump) starts near 300 meV and tracks the LDA dispersion. This two-branched character is reminiscent of the two poles solution for bosonic modes coupling considered in ref. 19, but the tiny intensity of the quasiparticle peak in the EDC indicates that the electron–phonon interaction is so strong here that the theory based on the conventional metallic state fails. In fact, denoting the quasiparticle effective mass by m^* obtained from the low-energy EDC dispersion and the bare band mass obtained from fitting the LDA dispersion by m_{LDA} , the value of the coupling constant λ expressed according to the Eliashberg textbook definition as $\lambda = [(m^*/m_{\text{LDA}}) - 1]$ is found to be ~ 4.6 (Fig. 2), and according to canonical theory, we would have expected the crystal to have run into a structural instability.

Much as in HTSCs, the Fermi surface maps show well-nested straight sections along the antinodal direction (Fig. 3). As for the

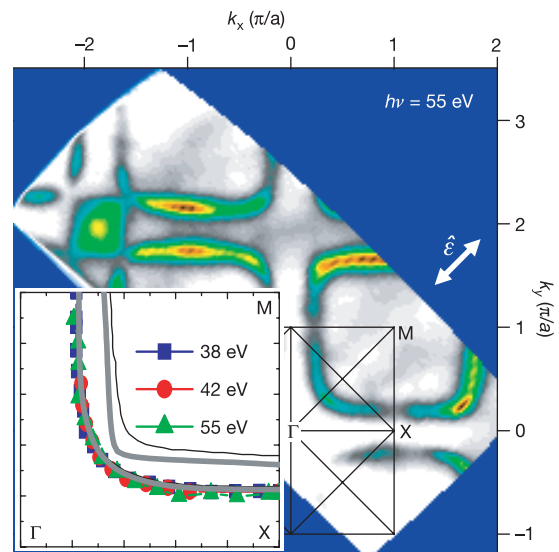


Figure 3 | Fermi surface topology. The Fermi surface has been determined by integrating the spectral weight in a ± 50 meV window around E_F of spectra excited with photons of 55 eV. Reducing the integration window to ± 10 meV results in the same spectral weight distribution. The momentum distribution of spectral weight is strongly affected by the photon energy and especially the experimental geometry, as a result of the impact that symmetry has on the matrix elements². According to LDA calculations, the Fermi surface topology consists of a small electron pocket derived from $3z^2$ -like states centred at the Γ point and $x^2 - y^2$ -derived hole pockets centred at the M point and split into two by the hybridization between the bilayers⁸ (Lin, H., Sahrakorpi, S., Barbiellini, B. & Bansil, A., personal communication). We found that (data not shown) the small electron pocket centred at Γ is enhanced when the photon polarization vector is perpendicular to the sample plane, unlike in the cases shown here, for which the light polarization vector lies in the sample plane directed along the nodal direction (white arrow). This geometry was nonetheless used because it gives better exposure of the spectral weight at E_F . To better quantify the Fermi surface, we made use of the MDC analysis, which identifies as Fermi wavevectors ($\mathbf{k}_F$) the maximum positions of the momentum distribution of spectral weight at E_F . The result is shown in the inset, where the Fermi surface contour extracted with the MDC analysis of data collected at three different photon energies is compared to the LDA calculations from ref. 8 (black line) and the work of Lin, H., Sahrakorpi, S., Barbiellini, B. & Bansil, A. (personal communication) (grey line). Despite an extensive search involving the use of different photon energy and/or experimental geometries, we observed only the $x^2 - y^2$ band closest to the Γ point, which tracks the LDA prediction extremely well, giving us confidence that our data are truly representative of the LSMO compound.

copper oxides, plots like those shown in Fig. 3 cannot discriminate between the true Fermi surface (sharp quasiparticle) and the ghost-like Fermi surface with no quasiparticle peaks. As indicated by the spectral lineshapes for different cuts through the Brillouin zone (Fig. 4), sharp quasiparticle peaks are only found close to the nodal points (Fig. 4a–c), while the spectra near the antinodal points (Fig. 4f) look quite similar to the spectra for corresponding momenta in the pseudogap state of underdoped HTSC or in a polaron state as exhibited in undoped insulating copper oxides such as $\text{Ca}_2\text{CuO}_2\text{Cl}_2$ and La_2CuO_4 (refs 15, 20). As in the copper oxides, the transition from well-defined nodal quasiparticle to the faint features of the segments along the antinodal directions takes place very rapidly in momentum space.

To illustrate the nodal–antinodal dichotomy further, we show in Fig. 4h and 4i the MDC spectra corresponding to the nodal and antinodal cuts denoted in Fig. 4j. In contrast to the nodal MDC spectra, which are sharply defined and dispersing, the antinodal MDC spectra are dispersionless for energies as high as 100 meV, a behaviour which is almost identical to that observed in $\text{Ca}_{2-x}\text{Na}_x\text{CuO}_2\text{Cl}_2$, further stressing the similarity to the nodal–antinodal dichotomy in the copper oxides¹². We searched extensively for the presence of quasiparticle peaks with different photon energies and experimental geometries, but did not find any evidence of a quasiparticle peak other than along the nodal direction within the resolution of our experiments, leading us to conclude that the quasiparticles along the straight Fermi surface segments in the antinodal regions are suppressed below the detection limit, with Fermi arcs in the vicinity of the nodal points where discernible and well-defined quasiparticles exist.

These observations provide an interpretation of the transport measurements. Relevant to the analysis of the transport parameters

is the strong mass enhancement of the quasiparticle, whose fitting yields a value for the effective mass $m^* \approx 3.3m_e \approx 5.6m_{\text{LDA}}$, where m_e denotes the free electron mass (Fig. 2). From the MDC peak width Δk at E_F we estimate the mean free path to be $l \approx 1/\Delta k = 25 \text{ \AA}$ which, when divided by the bare Fermi velocity²¹, yields a scattering rate of $\tau \approx 3.28 \text{ fs}$. Using these parameters in the Drude formula, we obtain an approximate value for the electrical resistivity $\rho = m^*/ne^2\tau \approx 0.66 \times 10^{-3} \Omega \text{ cm}$, in fairly good agreement with the in-plane resistivity value of $\sim 2 \times 10^{-3} \Omega \text{ cm}$ measured by transport²². This result indicates that the transport properties of this compound are determined by the nodal quasiparticles and that their low mobility $\mu = e\tau/m^*$ is responsible for the relatively high value (for a metal) of the resistivity. Explaining the coexistence of metallic transport with the ghost-like straight Fermi surface segments, the presence of the nodal quasiparticles is thus a key missing element that reconciles the metallic transport properties with previous ARPES measurements in the bilayer manganites.

Our findings demonstrate that the ferromagnetic-metallic ground state in the bilayer manganites is far from being the highly delocalized metal described by the textbook double-exchange model^{6,7}. Despite the immense differences between their electronic structures both at high energy (the large 1.8 eV ‘clean’ dispersion in the manganite versus the small 0.3–0.4 eV ‘bare’ bandwidth and strong damping in copper oxides due to antiferromagnetism) and at low energy (ferromagnetism versus superconductivity), underdoped copper oxides and layered manganites have a surprisingly similar mechanism in common that produces the characteristics of a nodal–antinodal dichotomous pseudogap metal. Metallic ferromagnetism and superconductivity are entirely different phenomena, so the characteristics of the pseudogap metal clearly do not have a direct connection with these ground-state orders, although the underlying physics may be

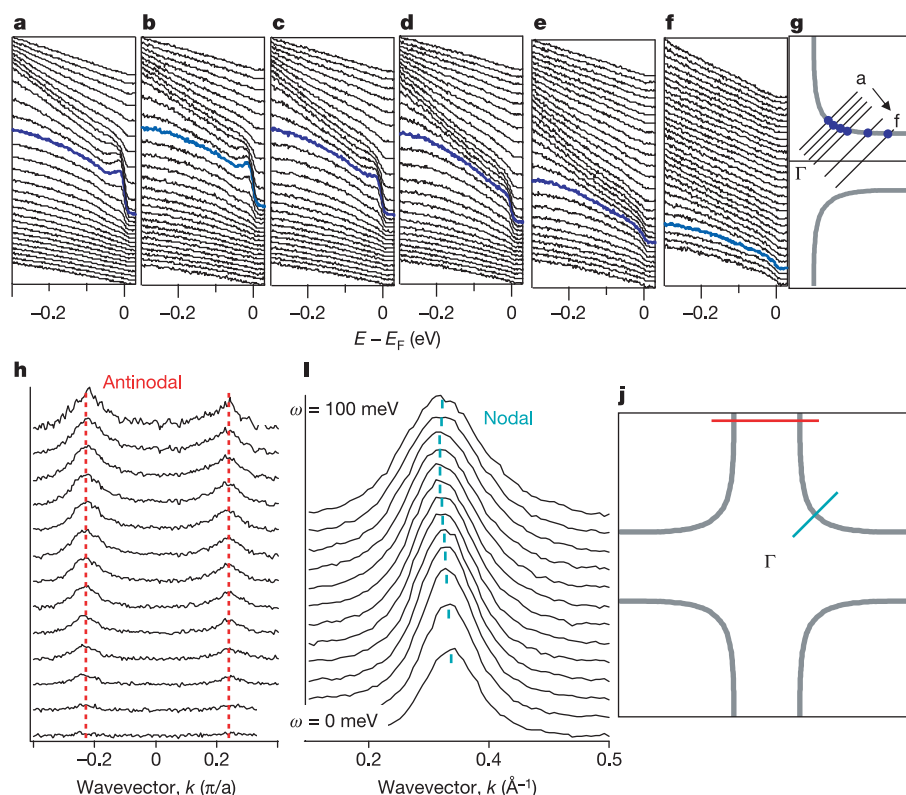


Figure 4 | Nodal-antinodal dichotomy in momentum space. a–f, EDC corresponding to the cuts shown in g. The spectral weight of the well-defined quasiparticle peak rapidly loses intensity while moving away from the nodal direction. h, MDC spectra for the antinodal; i, nodal cuts shown in j. The presence of quasiparticle peaks along the nodal direction and their absence

along the antinodal sections of the Fermi surface is not related to an anisotropic distribution of spectral intensity due to matrix element effects. This is evident from Fig. 3, which shows that there are regions in the parallel Fermi surface sections that contain strong spectral intensity, but do not yield well-defined quasiparticle peaks such as those along the node.

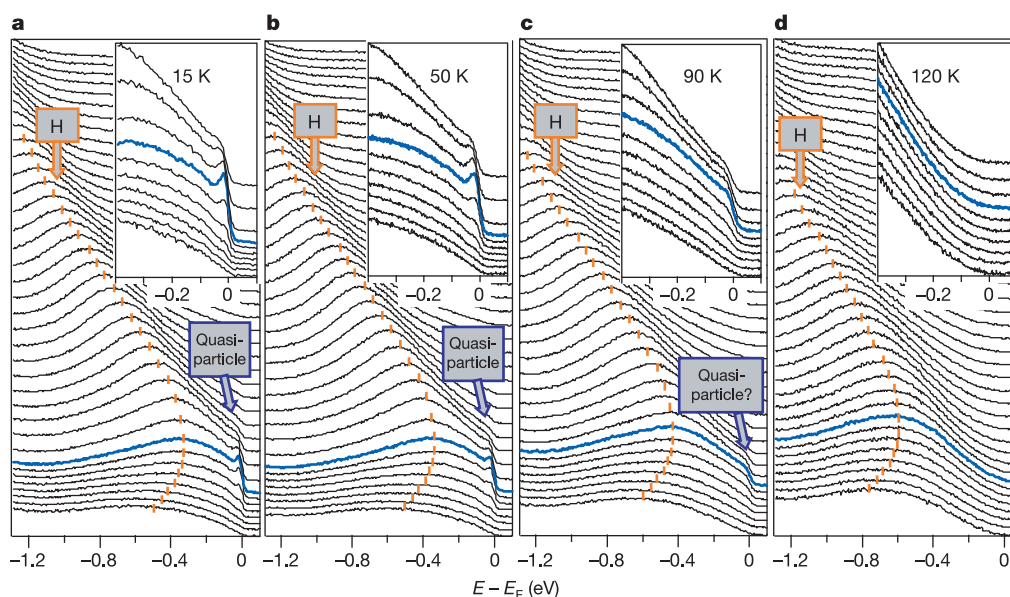


Figure 5 | Temperature-dependent evolution of the nodal quasiparticle. The data have been collected along the $(0,0)-(\pi,\pi)$ nodal direction with photon energy $h\nu = 42$ eV and light polarization vector lying in the sample plane. Note the very small changes with temperature of the broad peak (hump, H) as opposed to the dramatic changes of the quasiparticle peak (insets). The quasiparticle peak—visible in **a** and **b**—becomes hard to track at $T = 90$ K $= 0.75T_C$ (**c**) and is undetectable at the Curie temperature

$T_C = 120$ K (**d**), which is concomitant with a metal–insulator transition between a low-temperature ferromagnetic-metallic ground state and a high-temperature paramagnetic-insulating phase. To ensure reproducibility of the data and exclude the occurrence of any surface damage during the measurements, the spectra have been collected by both increasing and decreasing the temperature.

important for both cases. Instead, our findings indicate that the solution of the nodal–antinodal dichotomy puzzle lies elsewhere and suggest the occurrence of a phase which is ubiquitous in transition metal oxides and theoretically poorly understood: the polaronic metal with anisotropic band structure.

Although the issue is not settled for the copper oxides, there is no doubt that for the manganites the electron–phonon coupling is an essential microscopic ingredient. The essence of the CMR effect is that at a temperature that is a tiny fraction of the Fermi energy the system turns from a non-degenerate liquid into a degenerate (apparently ‘nodal’) Fermi liquid²³. Scattering measurements have provided direct evidence that the high-temperature classical liquid is formed from lattice polarons²⁴. These form in turn a strongly correlated liquid phase, showing pronounced short-range-ordered correlations²⁵. The onset of ferromagnetism tips the balance, via the double-exchange mechanism, in favour of a quantum-coherent, degenerate metal²⁶. Our data support the notion that on microscopic scales this low-temperature metal is not very different from the high-temperature liquid.

Moreover, the local charge-ordered correlations observed at T_C seem to have a precursor already at temperatures as low as 20 K ($T/T_C \approx 0.16$), discernible as the presence of the well-nested straight sections of the Fermi surface. For this reason, the ferromagnetic-metallic ground state may be referred to as an ‘incipient’ precursor to the charge-ordering state. The intriguing correlation between the nesting characteristics of the antinodal ghost-like Fermi surface and the observed charge ordering may not be coincidental. In fact, ARPES investigations of underdoped $\text{Ca}_{2-x}\text{Na}_x\text{CuO}_2\text{Cl}_2$, for which checkerboard charge ordering has been reported²⁷, revealed the existence of quasiparticles only along the nodal direction and the intriguing correlation between the charge-ordering periodicity and the antinodal parallel segments of the ghost-like Fermi surface¹². To confirm this additional commonality with the copper oxides will require a systematic study of the doping dependence of the Fermi surface and charge correlations.

Our findings offer some guidance for the development of a theoretical understanding of this mysterious metallic state. We

have already emphasized that the conventional metal theory (that is, the Migdal–Eliashberg theory) has non-physical ramifications when applied to the ‘self-energy kinks’ as shown in Figs 1 and 2. Instead, it makes more sense to invoke recent results from polaron theory^{16,17}. For a single polaron the spectral function consists of a low-energy ‘zero-phonon’ quasiparticle peak representing the centre of mass of the quantum motion of the polaron, and a high-energy incoherent resonance. It was recently demonstrated that in general the peak position of the incoherent resonance has to track closely the bare dispersion^{9, 15–17}. This reflects the motions of the undressed electron confined in the polaron cloud, and a similar picture has been suggested for underdoped copper oxides¹⁵.

In bilayer manganites, the temperature-dependent evolution of the ‘peak–dip–hump’ structure in the EDC (Fig. 5) offers an opportunity to confirm that this interpretation of the high-energy dispersion shown in Figs 1 and 2 is indeed correct. As expected within this interpretation, upon increasing temperature the high-energy branch does not undergo dramatic changes, while the quasiparticle peak disappears because the polarons lose the coherence associated with the polaron centre of mass. The challenge facing theory is to explain why the centre of mass sector (quasiparticle) turns itself into a ‘nodal’ Fermi liquid characterized by the Fermi arcs.

METHODS

Our experiments were performed on high-quality single crystals of composition $\text{La}_{1.2}\text{Sr}_{1.8}\text{Mn}_2\text{O}_7$ ($x = 0.4$, LSMO) grown with the floating-zone method²⁸. The ARPES measurements were carried out on beamline 10.0.1 at the Berkeley Advanced Light Source (ALS) using a Scienta R4000 electron analyser in the angle mode, which allows spectra to be recorded simultaneously within an angular aperture of 30° or 14° . We used the 30° mode for the Fermi surface maps and the 14° mode for the spectra with higher momentum resolution. The angular resolutions of the 30° and the 14° modes are $\pm 0.5^\circ$ and $\pm 0.15^\circ$, corresponding to $\sim 0.03 \text{ \AA}^{-1}$ and $\sim 8.3 \times 10^{-3} \text{ \AA}^{-1}$, respectively. The samples were cleaved in ultrahigh vacuum ($\sim 1\text{--}2 \times 10^{-11}$ torr) and measured at 20 K with a total instrumental energy resolution of 12–25 meV. The outermost 10-Å-thick bilayer in LSMO single crystals cleaved at room temperature (300 K) in air is insulating, with no long-range ferromagnetic order²⁹. Our findings of a

metallic ground state are not inconsistent with the results of ref. 29, owing to the different surface preparation methods.

Received 27 July; accepted 19 September 2005.

- Shen, Z.-X. *et al.* Anomalous large g anisotropy in the a - b plane of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **70**, 1553–1556 (1993).
- Damascelli, A., Shen, Z. X. & Hussain, Z. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).
- Emery, V. J. & Kivelson, S. A. Superconductivity in bad metals. *Phys. Rev. Lett.* **74**, 3253–3256 (1995).
- Honerkamp, C. & Lee, P. A. Staggered flux vortices and the superconducting transition in the layered cuprates. *Phys. Rev. Lett.* **92**, 177002 (2004).
- Chakravarty, S., Laughlin, R. B., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503 (2001).
- Tokura, Y. *Colossal Magnetoresistive Oxides* Ch. 1, 1 (Gordon & Breach Science, Amsterdam, 2000).
- Dagotto, E. *Nanoscale Phase Separation And Magnetoresistance* (series eds Cardona, M. *et al.*) Ch. 2, 13 (Springer, Berlin/Heidelberg/New York, 2002).
- Huang, X. Y., Mryasov, O. N., Novikov, D. L. & Freeman, A. J. Electronic and magnetic properties of layered colossal magnetoresistive oxides: $\text{La}_{1+2x}\text{Sr}_{2-2x}\text{Mn}_2\text{O}_7$. *Phys. Rev. B* **62**, 13318–13322 (2000).
- Dessau, D. S. *et al.* k -Dependent electronic structure, a large “ghost” Fermi surface, and a pseudogap in a layered magnetoresistive oxide. *Phys. Rev. Lett.* **81**, 192–195 (1998).
- Saitoh, T. *et al.* Temperature-dependent pseudogaps in colossal magnetoresistive oxides. *Phys. Rev. B* **62**, 1039–1043 (2000).
- Chuang, Y.-D., Gromko, A. D., Dessau, D. S., Kimura, T. & Tokura, Y. Fermi surface nesting and nanoscale fluctuating charge/orbital ordering in colossal magnetoresistive oxides. *Science* **292**, 1509–1513 (2001).
- Shen, K. M. *et al.* Nodal quasiparticles and antinodal charge ordering in $\text{Ca}_{2-x}\text{Na}_x\text{CuO}_2\text{Cl}_2$. *Science* **307**, 901–904 (2005).
- Wells, B. O. *et al.* E versus k relations and many body effects in the model insulating copper oxide $\text{Sr}_2\text{CuO}_2\text{Cl}_2$. *Phys. Rev. Lett.* **74**, 964–967 (1995).
- Ronning, F. *et al.* Anomalous high-energy dispersion in angle-resolved photoemission spectra from the insulating cuprate $\text{Ca}_2\text{CuO}_2\text{Cl}_2$. *Phys. Rev. B* **71**, 094518 (2005).
- Shen, K. M. *et al.* Missing quasiparticles and the chemical potential puzzle in the doping evolution of the cuprate superconductors. *Phys. Rev. Lett.* **93**, 267002 (2004).
- Mishchenko, A. S. & Nagaosa, N. Electron-phonon coupling and a polaron in the t - j model: from the weak to the strong coupling regime. *Phys. Rev. Lett.* **93**, 036402 (2004).
- Rosch, O. & Gunnarsson, O. Dispersion of incoherent spectral features in systems with strong electron-phonon coupling. *Eur. Phys. J. B* **43**, 11–18 (2005).
- Perebeinos, V. & Allen, P. B. Franck–Condon-broadened angle-resolved photoemission spectra predicted in LaMnO_3 . *Phys. Rev. Lett.* **85**, 5178–5181 (2000).
- Englesberg, S. & Schrieffer, J. R. Coupled electron-phonon system. *Phys. Rev.* **131**, 993–1008 (1963).
- Yoshida, T. *et al.* Metallic behaviour of lightly doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ with a Fermi surface forming an arc. *Phys. Rev. Lett.* **91**, 027001 (2003).
- Kaminski, A. *et al.* Renormalization of spectral line shape and dispersion below T_c in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **86**, 1070–1073 (2001).
- Moritomo, Y., Asamitsu, A., Kuwahara, H. & Tokura, Y. Giant magnetoresistance of manganese oxides with a layered perovskite structure. *Nature* **380**, 141–144 (1996).
- Palstra, T. T. M. *et al.* Transport mechanisms in doped LaMnO_3 : evidence for polaron formation. *Phys. Rev. B* **56**, 5104–5107 (1997).
- Vasiliu-Doloc, L. *et al.* Charge melting and polaron collapse in $\text{La}_{1.2}\text{Sr}_{1.8}\text{Mn}_2\text{O}_7$. *Phys. Rev. Lett.* **83**, 4393–4396 (1999).
- Campbell, B. J. *et al.* Structure of nanoscale polaron correlations in $\text{La}_{1.2}\text{Sr}_{1.8}\text{Mn}_2\text{O}_7$. *Phys. Rev. B* **65**, 014427 (2001).
- Millis, A. J., Littlewood, P. B. & Shraiman, B. I. Double exchange alone does not explain the resistivity of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **74**, 5144–5147 (1995).
- Hanaguri, T. *et al.* A ‘checkerboard’ electronic crystal state in lightly hole-doped $\text{Ca}_{2-x}\text{Na}_x\text{CuO}_2\text{Cl}_2$. *Nature* **430**, 1001–1005 (2005).
- Mitchell, J. F. *et al.* Charge delocalization and structural response in layered $\text{La}_{1.2}\text{Sr}_{1.8}\text{Mn}_2\text{O}_7$: Enhanced distortion in the metallic regime. *Phys. Rev. B* **55**, 63–66 (1997).
- Freeland, J. W. *et al.* Full bulk polarization and intrinsic tunnel barriers at the surface of layered manganites. *Nature Mater.* **4**, 62–67 (2005).

Acknowledgements The work at the ALS and SSRL is supported by the DOE Office of Basic Energy Science, Division of Material Science. The work at Stanford is also supported by an NSF grant and an ONR grant. The work at Argonne National Laboratory is supported by the US Department of Energy Office of Science.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.-X.S. (zxshen@stanford.edu) or N.M. (NMannella@lbl.gov).

Probing carrier dynamics in nanostructures by picosecond cathodoluminescence

M. Merano¹, S. Sonderegger¹, A. Crottini¹, S. Collin¹, P. Renucci¹, E. Pelucchi¹, A. Malko¹, M. H. Baier¹, E. Kapon¹, B. Deveaud¹ & J.-D. Ganière¹

Picosecond and femtosecond spectroscopy allow the detailed study of carrier dynamics in nanostructured materials¹. In such experiments, a laser pulse normally excites several nanostructures at once. However, spectroscopic information may also be acquired using pulses from an electron beam in a modern electron microscope, exploiting a phenomenon called cathodoluminescence. This approach offers several advantages. The multimode imaging capabilities of the electron microscope enable the correlation of optical properties (via cathodoluminescence) with surface morphology (secondary electron mode) at the nanometre scale². The broad energy range of the electrons can excite wide-bandgap materials, such as diamond- or gallium-nitride-based structures that are not easily excited by conventional optical means. But perhaps most intriguingly, the small beam can probe a single selected nanostructure. Here we apply an original time-resolved cathodoluminescence set-up to describe carrier dynamics within single gallium-arsenide-based pyramidal nanostructures³ with a time resolution of 10 picoseconds and a spatial resolution of 50 nanometres. The behaviour of such charge carriers could be useful for evaluating elementary components in quantum computers^{4,5}, optical quantum gates⁶ or single photon sources^{7–9} for quantum cryptography¹⁰.

It is of primary importance to understand as precisely as possible how the charge carriers behave in a single nano-object^{11,12}. Standard optical techniques, such as fluorescence microscopy, provide a spatial

resolution that remains limited to a fraction of a micrometre. More sophisticated techniques, like near-field optical microscopy, can be used to gain spectroscopic information at the 100-nanometre scale, but do not routinely provide adequate time resolution¹³ when faint secondary signals, such as luminescence, have to be measured. Moreover, the performances of ultraviolet optics limit the use of these techniques to materials with a gap lower than 4 eV.

Modern electron microscopes represent a direct approach for determining surface morphology of semiconducting nanostructures with a spatial resolution of a few nanometres. Using the cathodoluminescence imaging mode, that is, an excitation of the luminescence via the electron beam of the microscope, we gain access to spectroscopic information within a single nano-object and correlation with the surface morphology is straightforward¹⁴. More than 20 years ago, Steckenborn *et al.*¹⁵ proposed a new set-up allowing the addition of temporal resolution to cathodoluminescence. Beam blanking was achieved by deflecting the electron beam away from the axis of the electron optical column, using a set of deflection plates. Even if this experimental technique made it possible to obtain interesting results^{16–20}, it suffers from some disadvantages: the beam always sweeps the sample and the minimum pulse length is a few tens of nanoseconds. The semiconductor is then excited in steady state and the temporal information is obtained during the decay of the luminescence only, with a temporal resolution not better than a few hundred picoseconds.

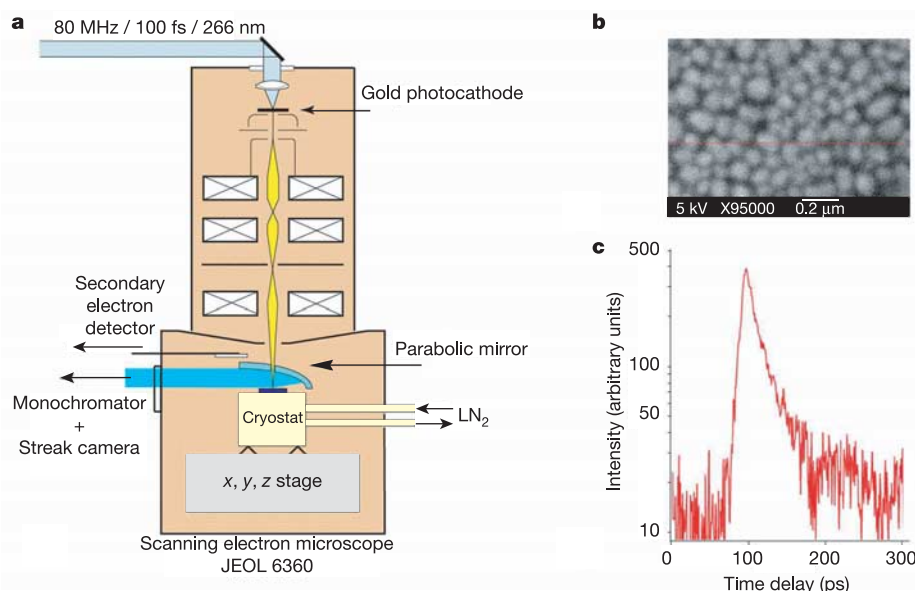


Figure 1 | Set-up, and spatial and temporal resolution. **a**, Schematic implementation of our time-resolved cathodoluminescence system. **b**, Secondary electron images of gold nanoparticles deposited on an amorphous carbon film. The spatial resolution on such a specimen is estimated to be better than 50 nm. The spatial resolution in secondary electron mode is mainly determined by the size of the electron probe onto the specimen. **c**, Time-resolved spectra of a GaN specimen at room temperature. From the rise time of the signal, the temporal width of the electron pulse is estimated to be about 10 ps.

¹Institute of Quantum Electronics and Photonics, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne-EPFL, Switzerland.

To overcome these drawbacks, we developed a new cathodoluminescence set-up, allowing us to keep the main advantages of the scanning electron microscope, that is, the multi-imaging mode capabilities, and to improve the pulsed electron gun. We use an optically driven electron gun^{21–24} similar to the one used to probe chemical changes by ultrafast electron diffraction²⁵.

To implement our idea we have replaced the classical hairpin tungsten electron gun of a JEOL 6360 scanning electron microscope (Fig. 1a) with a 20-nm-thick gold photocathode deposited on a quartz plate. By illuminating the photocathode (in transmission mode) by a ultraviolet mode-locked laser (266 nm, 200 fs, 80.7 MHz of repetition rate, spot size radius of 3 μm), we generate a pulsed electron beam, bright enough (average brightness of $2 \times 10^3 \text{ A cm}^{-2} \text{ steradians}^{-1}$, 1–100 pA at the specimen surface) to allow secondary electron images to be recorded, thus allowing us to identify the nanostructures under study. A streak camera (Hamamatsu 5680), mounted behind a monochromator (spectral resolution better than 200 μeV) is used as a time-resolved detector (in photon-counting mode). A liquid-nitrogen cryostat is used to cool the sample down to 90 K. Observations on a specimen made of gold particles deposited on an amorphous carbon film confirmed that this modified microscope has a spatial resolution of the secondary electron images of better than 50 nm (Fig. 1b). Careful measurements on a reference GaN specimen demonstrate an electron pulse width of about 10 ps (Fig. 1c). The average number of electrons per pulse is generally lower than ten (a beam current of 13 pA corresponds to one electron per pulse) still allowing us to realize secondary electron images of the specimen. We note here that a single electron per pulse still allows a large signal because, depending on the acceleration voltage, each electron may generate a few thousand electron-hole pairs into the sample. To estimate the potentiality of this new set-up, we choose to investigate GaAs-based pyramidal nanostructures.

Our choice is dictated by the interest of such structures both in the field of basic research and for a number of possible applications. Pyramidal quantum dots (QDs) naturally provide site control, a wide choice in the possible densities, excellent optical properties, high uniformities and flexible energy tuning^{3,9,26}. The simultaneous presence of several nanostructures in one single pyramid has attracted the interest of innovative theoretical proposals for quantum information protocols²⁷. The samples were fabricated with the following

procedure: before growth, a GaAs substrate is patterned with a 5- μm pitch hexagonal matrix of tetrahedral recesses, using standard photolithography and wet etching³. Growth of InGaAs/AlGaAs heterostructures by low-pressure organo-metallic chemical vapour deposition is then performed on the patterned substrate. The growth process results in the formation of different nanostructures, as sketched and defined in Fig. 2.

All measurements have been performed at 90 K and an acceleration voltage of 10 kV. The cathodoluminescence spectrum from a single pyramid (obtained with a pulsed excitation and with a time-integrated detection) and spectrally resolved images taken at wavelengths corresponding to the different recombination channels are presented in Fig. 3. Some of the peaks in the spectrum were identified on the basis of previous optical studies, through their energy position and the cathodoluminescence image²⁸.

Importantly, the cathodoluminescence images relative to QDs, quantum wires (QWRs), and vertical quantum wires (VQWRs) (Fig. 2) cannot be directly related to the nanostructure location at the origin of the luminescence signal. This is typical for cathodoluminescence experiments because the emission pattern does not depend only on the primary electron diffusion range, but also on the carrier transport processes in these complex nanostructures. If we consider, for example, the cathodoluminescence image at the QD wavelength, we observe that the emission pattern spans the entire pyramid. This means that carriers created at any point of the pyramid find a path to diffuse towards the dot. One of the main advantages of time-resolved cathodoluminescence is that it allows us to unveil the specificities of this carrier transport. We can excite at any point of interest and gain information in the spectral, spatial and temporal domains. As an example, results obtained when we excited at the centre of one of the pyramid faces are presented in Fig. 4.

To give more quantitative estimates on the carrier diffusion properties within the structure, we have excited the pyramid at carefully chosen points (three along an edge and three points onto the face of the pyramid) and we have detected in parallel the signals corresponding to the different recombination channels (see Fig. 2): QD, VQWR, QWR, quantum well (QW) and vertical quantum well (VQW). Typical results for the VQW and the QWR are depicted in

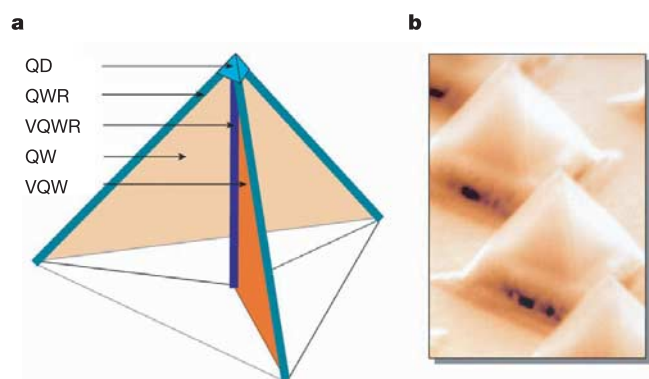


Figure 2 | InGaAs/AlGaAs pyramidal quantum structures. **a**, Schematic representation of the five heterostructures present in an InGaAs/AlGaAs pyramid (the AlGaAs cap layer is not represented on the scheme). An InGaAs quantum dot (QD; typically 30 nm in lateral size) connected to several types of low-dimensional barriers: three InGaAs quantum wires (QWRs) on the edges of the pyramid, three InGaAs quantum wells (QWs) on the pyramid faces, and three vertical quantum wells (VQWs) and a vertical quantum wire (VQWR) located at the intersection of the vertical quantum wells. **b**, Scanning electron microscope picture of the pyramid (the top of the pyramids are separated by 5 μm). The final step of the process is a back etching to remove the substrate.

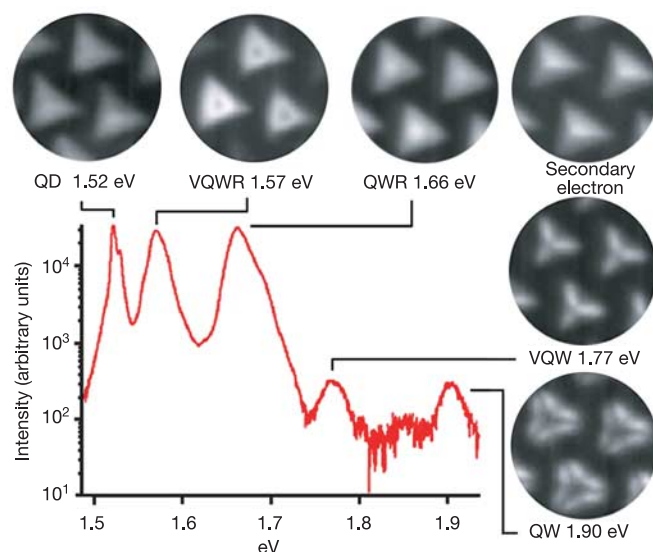


Figure 3 | Cathodoluminescence spectrum and images. Cathodoluminescence characterization of our sample. The spectrum shows five main peaks corresponding to the five main structures in the pyramid. These are, from low to high energy, the quantum dot at the tip of the pyramid (QD), the vertical quantum wire (VQWR), the lateral quantum wires (QWR) the vertical quantum wells (VQW) and the quantum well along the sides of the pyramid (QW).

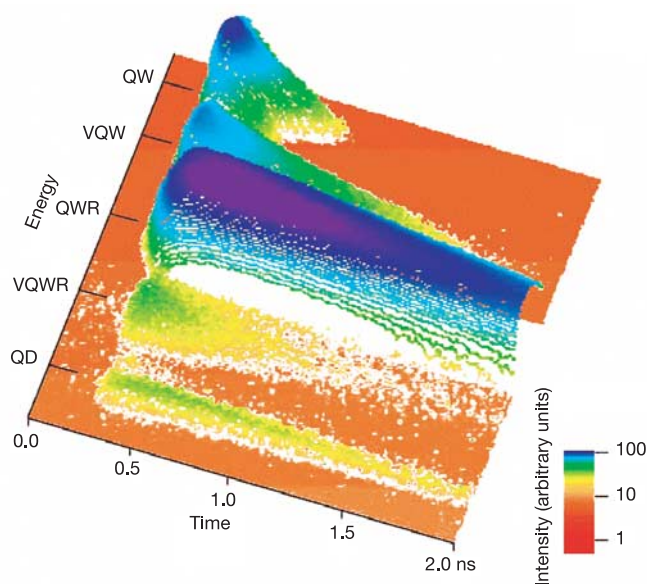


Figure 4 | Streak image. Typical time traces corresponding to the different nanostructures of the pyramid, when excited in the middle of one of the faces of this pyramid. The different curves are labelled accordingly to the related structure. The five different nanostructures display very different time behaviours that are signatures of the complex transport mechanisms within the pyramid.

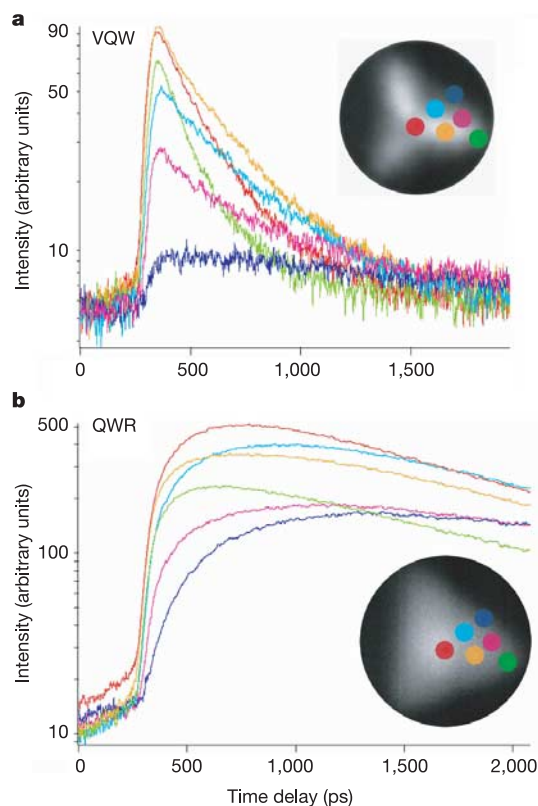


Figure 5 | Luminescence temporal profiles versus excitation points. Time traces corresponding to the VQW (a) and the QWR (b) for the different excitation points depicted in the inset. From such traces we can directly deduce that: (1) the carrier injection in the VQW and the QWR is more efficient when we excite on the pyramid edge; (2) the rise time for the QWR basically corresponds to the decay time of the VQW; and (3) long delays are observed when excited on the side of the pyramid resulting from the longer diffusion path.

Fig. 5 for six different excitation points within the pyramid. The sensitivity of the method is directly demonstrated by the large differences observed between the different traces, both in terms of intensities and in terms of dynamics.

Full analysis of the data allows us to obtain a very detailed understanding of the carrier dynamics inside these pyramids. The electron–hole pairs are created within a volume of 300 nm around the impact point of the electrons, that is, mostly in the AlGaAs barriers of the structure. From there, they diffuse to the closest nanostructure, which is often one of the VQWs or the QWRs. This process occurs through diffusion, with a diffusivity of about $10 \text{ cm}^2 \text{ s}^{-1}$ corresponding to an ambipolar mobility of the order of $1,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, and to a trapping time shorter than our experimental resolution. This explains in particular the 150-ps rise time observed for the QWR luminescence. The carriers in the VQW get trapped in the QWR within a typical time of less than 100 ps, and then diffuse towards the central structures (the VQWR and the QD) in the QWR. We show in Fig. 6 the results of our diffusion model for three different excitation points along the edge of the pyramid, allowing us to limit the numbers of parameters for the fit. The model allows us to deduce a mobility of $1,400 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ along the wire, and a negligible quantum mechanical trapping time into the VQWR. In the case of Figs 5 and 6, owing to the high mobility in the selected pyramid, saturation effects are observed both in the VQWR and in the QD. Indeed, even if we excite at the level of one electron per pulse, this electron creates enough pairs in the structure to allow this because the QD can only accommodate two pairs, and the VQWR about ten pairs.

Thus we have developed a new spectroscopic tool, a picosecond time-resolved cathodoluminescence system with a time resolution of

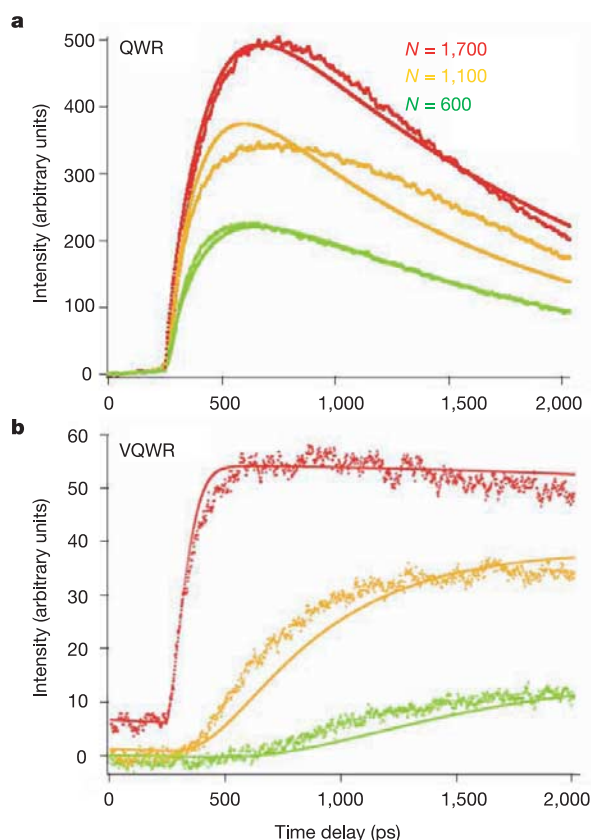


Figure 6 | QWR and VQWR fits. Fits to the time traces of the QWR (a) and the VQWR (b), with basically one single fitting parameter: the ambipolar diffusion coefficient of the carriers within the QWR. Saturation of the VQWR as well as the very long radiative and non-radiative lifetimes is clearly shown on the upper curve of Fig. 6.

the order of 10 ps, together with a spatial resolution of 50 nm. This system allows us to study the carrier dynamics inside GaAs back-etched pyramids, and obtain a precise knowledge of the photoexcited carrier path towards the lower-energy state: the dot at the tip of the pyramid.

Received 2 February; accepted 15 September 2005.

- Shah, J. *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* Ch. 8 (Springer, Berlin, 1999).
- Reimer, L. *Scanning Electron Microscopy* Ch. 1 (Springer, Berlin, 1998).
- Kapon, E. *et al.* Site- and energy-controlled pyramidal quantum dot heterostructures. *Physica E* **25**, 288–297 (2004).
- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Imamoglu, A. Are quantum dots useful for quantum computation? *Physica E* **16**, 47–50 (2003).
- Li, X. *et al.* An all-optical quantum gate in a semiconductor quantum dot. *Science* **301**, 809–811 (2003).
- Benson, O., Santori, C., Pelton, M. & Yamamoto, Y. Regulated and entangled photons from a single quantum dot. *Phys. Rev. Lett.* **84**, 2513–2516 (2000).
- Gerard, J. M. & Gayral, B. Strong Purcell effect for InAs quantum boxes in three-dimensional solid-state microcavities. *J. Lightwave Technol.* **17**, 2089–2095 (1999).
- Baier, M. H. *et al.* Single photon emission from site-controlled pyramidal quantum dots. *Appl. Phys. Lett.* **84**, 648–650 (2004).
- Gisin, N., Ribordy, G. G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
- Warburton, R. J. Self-assembled semiconductor quantum dots. *Contemp. Phys.* **43**, 351–364 (2002).
- Korovyanko, O. J., Sheng, C. X., Vardeny, Z. V., Dalton, A. B. & Baughman, R. H. Ultrafast spectroscopy of excitons in single-walled carbon nanotubes. *Phys. Rev. Lett.* **92**, 017403 (2004).
- Gustafsson, A., Pistol, M. E., Montelius, L. & Samuelson, L. Local probe techniques for luminescence studies of low-dimensional semiconductor structures. *J. Appl. Phys.* **84**, 1715–1775 (1998).
- Menniger, J., Jahn, U., Brandt, O., Yang, H. & Ploog, K. Identification of optical transitions in cubic and hexagonal GaN by spatially resolved cathodoluminescence. *Phys. Rev. B* **53**, 1881–1885 (1996).
- Steckenborn, A., Munzel, H. & Bimberg, D. Cathodoluminescence lifetime pattern of GaAs-surfaces around dislocations. *J. Lumin.* **24–5**, 351–354 (1981).
- Bell, A. *et al.* Exciton freeze-out and thermally activated relaxation at local potential fluctuations in thick $\text{Al}_x\text{Ga}_{1-x}\text{N}$ layers. *J. Appl. Phys.* **95**, 4670–4674 (2004).
- Vetter, U., Hofsass, H. & Taniguchi, T. Visible cathodoluminescence from Eu-implanted single- and polycrystal c-BN annealed under high-temperature, high-pressure conditions. *Appl. Phys. Lett.* **84**, 4286–4288 (2004).
- Diaz-Guerra, C., Piqueras, J., Castaldini, A., Cavallini, A. & Polenta, L. Time-resolved cathodoluminescence and photocurrent study of the yellow band in Si-doped GaN. *J. Appl. Phys.* **94**, 2341–2346 (2003).
- Turck, V. *et al.* Time-resolved spectroscopy of single quantum dots: Evidence for phonon-assisted carrier feeding. *Phys. Status Solidi B* **224**, 643–647 (2001).
- Zhang, X., Rich, D. H., Kobayashi, J. T., Kobayashi, N. P. & Dapkus, P. D. Carrier relaxation and recombination in an InGaN/GaN quantum well probed with time-resolved cathodoluminescence. *Appl. Phys. Lett.* **73**, 1430–1432 (1998).
- Elsayedali, H. E. & Herman, J. W. Ultrahigh-vacuum picosecond laser-driven electron-diffraction system. *Rev. Sci. Instrum.* **61**, 1636–1647 (1990).
- May, P., Halbout, J. M. & Chiu, G. Picosecond photoelectron scanning electron-microscope for noncontact testing of integrated-circuits. *Appl. Phys. Lett.* **51**, 145–147 (1987).
- Kong, S. H., Kinrosswright, J., Nguyen, D. C., Sheffield, R. L. & Weber, M. E. Performance of cesium telluride photocathodes as an electron source for the Los-Alamos advanced fel. *Nucl. Instrum. Methods A* **358**, 284–286 (1995).
- Siwick, B. J., Dwyer, J. R., Jordan, R. E. & Miller, R. J. D. An atomic-level view of melting using femtosecond electron diffraction. *Science* **302**, 1382–1385 (2003).
- Williamson, J. C., Cao, J. M., Ihee, H., Frey, H. & Zewail, A. H. Clocking transient chemical changes by ultrafast electron diffraction. *Nature* **386**, 159–162 (1997).
- Hartmann, A., Ducommun, Y., Kapon, E., Hohenester, U. & Molinari, E. Few-particle effects in semiconductor quantum dots: Observation of multicharged excitons. *Phys. Rev. Lett.* **84**, 5648–5651 (2000).
- Piermarocchi, C., Chen, P., Sham, L. J. & Steel, D. G. Optical RKKY interaction between charged semiconductor quantum dots. *Phys. Rev. Lett.* **89**, 167402 (2002).
- Leifer, K., Hartmann, A., Ducommun, Y. & Kapon, E. Carrier transport and luminescence in inverted-pyramid quantum structures. *Appl. Phys. Lett.* **77**, 3923–3925 (2000).

Acknowledgements We thank N. Hoan and J. P. Hervé from OPEA and our technicians R. Roachat and N. Leiser for their help. We are also indebted to K. Leifer for scientific discussions of our results. We also thank M. Gatri and F. Füzesi, who developed part of the set-up for their diploma. This work was supported by the Swiss National Science Foundation, NCCR project 'Quantum Photonics'.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (michele.merano@ensta.fr).

Impacts of orbital forcing and atmospheric carbon dioxide on Miocene ice-sheet expansion

Ann Holbourn¹, Wolfgang Kuhnt¹, Michael Schulz² & Helmut Erlenkeuser³

The processes causing the middle Miocene global cooling, which marked the Earth's final transition into an 'icehouse' climate about 13.9 million years ago (Myr ago)^{1–4}, remain enigmatic. Tectonically driven circulation changes^{5,6} and variations in atmospheric carbon dioxide levels^{7,8} have been suggested as driving mechanisms, but the lack of adequately preserved sedimentary successions has made rigorous testing of these hypotheses difficult. Here we present high-resolution climate proxy records, covering the period from 14.7 to 12.7 million years ago, from two complete sediment cores from the northwest and southeast subtropical Pacific Ocean. Using new chronologies through the correlation to the latest orbital model⁹, we find relatively constant, low summer insolation over Antarctica coincident with declining atmospheric carbon dioxide levels at the time of Antarctic ice-sheet expansion and global cooling, suggesting a causal link. We surmise that the thermal isolation of Antarctica played a role in providing sustained long-term climatic boundary conditions propitious for ice-sheet formation. Our data document that Antarctic glaciation was rapid, taking place within two obliquity cycles, and coincided with a striking transition from obliquity to eccentricity as the drivers of climatic change.

About 13.9 Myr ago, the Earth's climate cooled dramatically after an extended period of relative warmth. This key transition in the Earth's climatic and biotic evolution marked the final stage of stepwise Cenozoic cooling, and coincided with major ice-sheet expansion over Antarctica^{1–4}. However, the timing, duration and driving mechanisms of this critical episode still remain largely unsolved, because sedimentary successions spanning this interval have been strongly affected by carbonate dissolution or burial diagenesis or proved incomplete owing to radical changes in ocean circulation. Thus, developing uninterrupted, high-resolution middle Miocene climate records that can be tied to an orbital timescale represents a major challenge for understanding the pacing and trigger of Cenozoic global cooling.

Here we present benthic foraminiferal oxygen ($\delta^{18}\text{O}$) and carbon ($\delta^{13}\text{C}$) isotope (4–5 kyr resolution) and X-ray fluorescence scanning (XRF) records (1-kyr resolution) in complete middle Miocene sedimentary sequences recovered at Ocean Drilling Program (ODP) Site 1146 (19° 27.40' N, 116° 16.37' E, 2,092 m) and Site 1237 (16° 0.421' S, 76° 22.685' W, 3,212 m) (Supplementary Information). These continuous successions allow unprecedented resolution of climate evolution on orbital timescales and permit detailed correlation of palaeoceanographic events in spatially separated regions in the northwest and southeast subtropical Pacific.

We developed orbitally tuned age models (Supplementary Information) by correlating the $\delta^{18}\text{O}$ records at each site to computed variations of the Earth's orbit and solar insolation (obliquity and eccentricity in ref. 9). The most salient features of the $\delta^{18}\text{O}$ time series

are the major increase in values at 13.91–13.84 Myr ago (0.8‰ at Site 1237 and 1.2‰ at Site 1146), and the transition from high-amplitude obliquity-paced variations dominant between 14.7 and 13.9 Myr ago to eccentricity-paced fluctuations between 13.8 and 13.1 Myr ago (Figs 1 and 2). Higher-amplitude oscillations at Site 1146 ($0.56\text{‰} \pm 0.12\text{‰}$ versus $0.31\text{‰} \pm 0.15\text{‰}$ mean amplitude at Site 1237) indicate a stronger temperature component in the 1146 $\delta^{18}\text{O}$ record between 14.7 and 13.9 Myr ago. Nevertheless, the global correlation of isotopic events and evidence for substantial Antarctic ice growth between 14.8 and 13.6 Myr ago¹⁰ support that both $\delta^{18}\text{O}$ records closely track ice-volume fluctuations. Furthermore, benthic Mg/Ca temperature estimates indicate that most of the deep-sea $\delta^{18}\text{O}$ increase at 13.91–13.84 Myr ago can be attributed to an increase in continental ice volume¹¹. The 100-kyr and 41-kyr signal components in $\delta^{18}\text{O}$ show amplitude modulation similar to that of the eccentricity and obliquity, with a striking transition from high amplitude in the 41-kyr band to high amplitude in the 100-kyr band at ~14.1–13.8 Myr ago (Fig. 2). The 100-kyr beat is also prominent in the 1237 Fe record after 14.1 Myr ago.

At first sight, the middle Miocene transition from a 41-kyr to 100-kyr world shows similarity to the 'Middle Pleistocene revolution' following an increase in Northern Hemisphere glaciation^{12–14}. However, the middle Miocene transition differs in several aspects: (1) the middle Miocene 100-kyr cycles do not exhibit higher-amplitude variations than the 41-kyr cycles; (2) the transition from 41-kyr to 100-kyr periodicity occurred during or even before ice-sheet expansion without substantial time lag; (3) the middle Miocene 100-kyr cycles show a clear phase relationship and similar amplitude modulation to eccentricity, as well as both 125-kyr and 95-kyr components (Fig. 1, Supplementary Fig. S4). In contrast, the origin of the middle Pleistocene 100-kyr cycles remains puzzling, because their onset coincided with a period of weak eccentricity forcing¹⁵.

Between 14.7 and 13.9 Myr ago, high amplitude in the 41-kyr band of the $\delta^{18}\text{O}$ signal (Fig. 2) suggests an already expanded Antarctic ice-sheet fluctuating in response to obliquity forcing. The ensuing $\delta^{18}\text{O}$ increase at 13.91–13.84 Myr ago, which marked the main increase in middle Miocene Antarctic glaciation, was also clearly modulated by obliquity (Figs 1 and 3). Obliquity controls the gradient in summer insolation between high and low latitudes, which drives the poleward atmospheric heat and moisture transfer¹⁶. Enhanced meridional moisture transport during periods of high differential heating between high and low latitudes at obliquity minima would promote ice-sheet growth, whereas a low summer insolation gradient at obliquity maxima would decrease poleward moisture transport, inhibiting ice-sheet build-up^{16,17}. Similar obliquity-paced variations in moisture supply for the Holocene and late Pleistocene East Antarctic ice-sheet are evident from the deuterium excess record in the Vostok ice core¹⁸. High deuterium excess values during obliquity

¹Institute of Geosciences, Christian-Albrechts-University, D-24118 Kiel, Germany. ²Department of Geosciences and Research Center Ocean Margins, University of Bremen, Postfach 330 440, D-28334 Bremen, Germany. ³Leibniz Laboratory for Radiometric Dating and Stable Isotope Research, Christian-Albrechts-University, D-24118 Kiel, Germany.

minima were related to an enhanced contribution of low latitude warm moisture sources leading to ice build-up, while during obliquity maxima cooler, local oceanic moisture sources dominated. In short, the prominent obliquity beat in benthic $\delta^{18}\text{O}$ before and during the major middle-Miocene ice-sheet expansion suggests that Southern Hemisphere ice growth was most probably primarily modulated by atmospheric heat and moisture transport, rather than by changing oceanic circulation patterns. On the other hand, ice-sheet dynamics may have in turn influenced oceanic circulation and heat transport. For instance, palaeocurrent reconstructions of the deep western boundary current reveal variability of the deep Pacific inflow on the 41-kyr timescale, with the highest current intensities during obliquity minima in the Pleistocene¹⁹ and middle Miocene before 13.9 Myr ago²⁰.

Stepwise cooling of Antarctic Circumpolar surface waters before the expansion of the cryosphere was inferred from high-latitude Mg/Ca temperature records from the South Tasman Rise ODP Site 1171 (ref. 21). When the Site 1171 benthic $\delta^{18}\text{O}$ is correlated to the orbitally tuned records of Sites 1146 and 1237, sea surface temperature (SST) variations between 14.20 and 13.87 Myr ago appear strongly influenced by orbital forcing (Fig. 3). A remarkable feature of the Site 1171 records is that opposite trends between ice volume and SSTs occur between 14.20 and 13.87 Myr ago: cooling in Antarctic Circumpolar surface waters coincided overall with ice volume decreases (lower benthic $\delta^{18}\text{O}$) at increasing obliquity,

whereas surface warming occurred during phases of ice build-up at decreasing obliquity (Fig. 3). These fluctuations are consistent with the notion that changing atmospheric heat and moisture supply drove ice volume on an obliquity timescale.

Between 14.05 and 13.87 Myr ago, three major cooling episodes occurred during an interval, when the amplitude of eccentricity increased (Fig. 3). This stepwise surface cooling was interpreted as eccentricity-paced intensification of the Antarctic Circumpolar Current, eventually leading to increased thermal isolation of Antarctica and global cooling²¹. However, it is intriguing that at least two of the three cooling episodes (centred at 14.04 and 13.88 Myr ago) coincided with decreasing benthic $\delta^{18}\text{O}$ (Fig. 3) and freshening of surface waters at Site 1171 (ref. 21). A fitting explanation may be that surface cooling was triggered by massive melting pulses, when the Antarctic ice sheet became unstable owing to the conjunction of increasing obliquity and high eccentricity favouring ice-sheet decay. The interval between 14.04 and 13.88 Myr ago was characterized by high-amplitude changes in insolation over Antarctica, and fluctuations in obliquity-driven heat and moisture supply would have led to ice-sheet instability and affected oceanic circulation. High-resolution deep-water temperature reconstructions to deconvolve ice volume and temperature in benthic $\delta^{18}\text{O}$ could potentially resolve the dynamics of Antarctic glaciation and its link to thermohaline circulation during this critical interval.

The main $\delta^{18}\text{O}$ increase after 13.9 Myr ago occurred during a

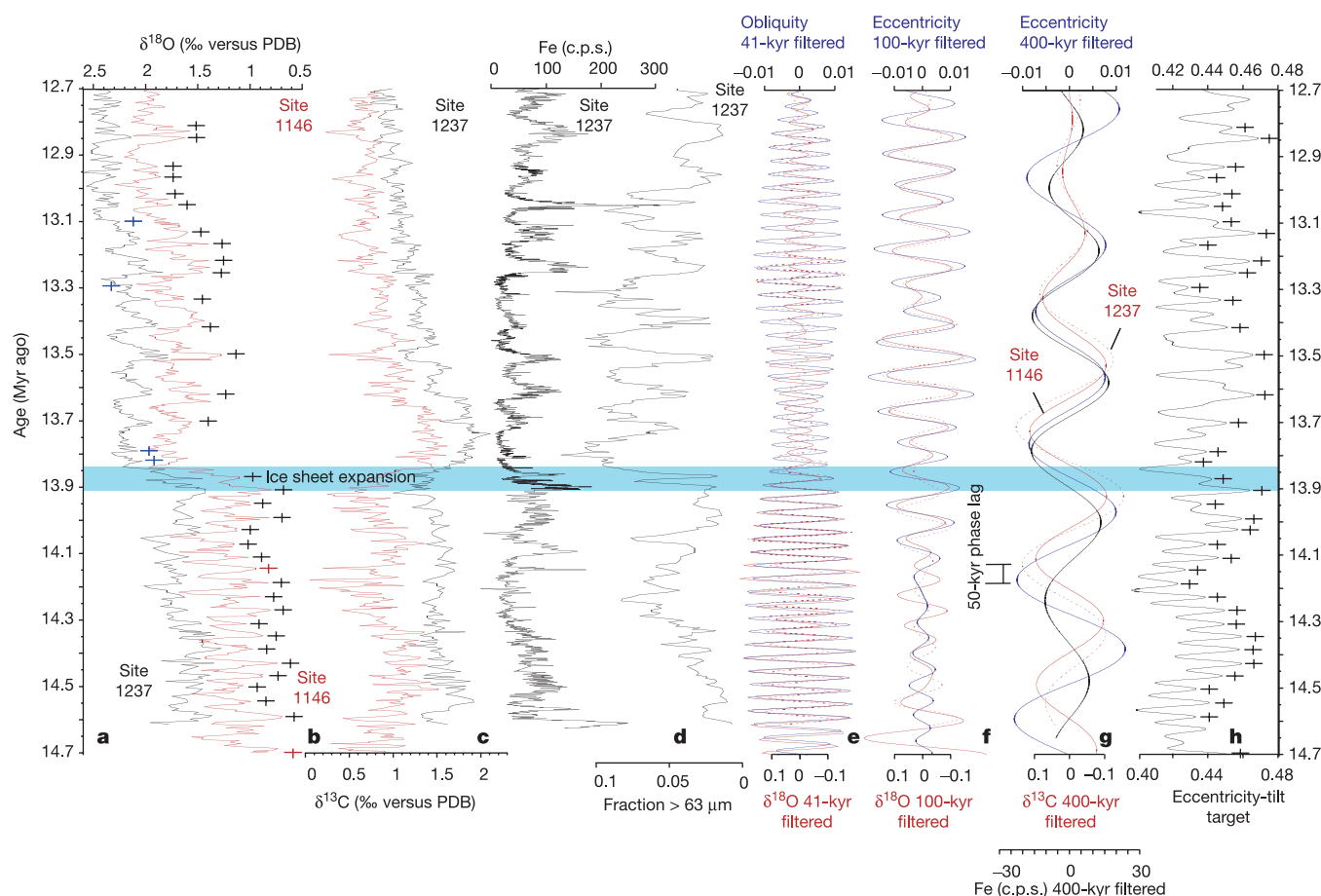


Figure 1 | Palaeoceanographic records from ODP Sites 1146 and 1237 showing that Earth's orbital configuration was the prime pacer of middle Miocene climate change. a, Benthic foraminiferal $\delta^{18}\text{O}$. Age control points marked by crosses (black, both sites; blue, Site 1237; red, Site 1146). b, Benthic foraminiferal $\delta^{13}\text{C}$. c, Site 1237 iron content in counts per second (c.p.s.). d, Site 1237 proportion of coarse fraction indicating increased

carbonate preservation during colder periods corresponding to low eccentricity after 13.9 Myr ago. e, f, Comparison of filtered 41-kyr and 100-kyr $\delta^{18}\text{O}$, obliquity⁹ and eccentricity⁹ (Site 1146, solid red line; Site 1237, dotted red line). g, Comparison of 400-kyr-filtered iron contents, $\delta^{13}\text{C}$ and eccentricity⁹. h, Eccentricity-tilt tuning target. Age correlation points marked by crosses.

period when eccentricity declined and amplitude variations in obliquity decreased (Fig. 3). This orbital configuration would have resulted in an extended period of low seasonal contrast over Antarctica, inhibiting summer melting and favouring ice-sheet expansion. Such an orbital configuration is not unique, and recurs approximately every 2.4 million years without inevitably inducing major ice growth. A further factor influencing Antarctic glaciation after 13.9 Myr ago may have been a coeval decrease in atmospheric CO_2 partial pressure (p_{CO_2}). Marine $\delta^{13}\text{C}$ is closely linked to p_{CO_2} through productivity and organic carbon burial feedbacks resulting in the lowest p_{CO_2} at $\delta^{13}\text{C}$ maxima and the highest p_{CO_2} at $\delta^{13}\text{C}$ minima²². A prominent feature of the middle Miocene $\delta^{13}\text{C}$ curve is the positive excursion starting at 13.92 Myr ago with the maximum centred at ~13.7 Myr ago, which represents the final and most pronounced $\delta^{13}\text{C}$ increase and p_{CO_2} decrease within the 'Monterey Excursion'⁸.

Enhanced carbonate preservation after 13.9 Myr ago could have contributed to increased organic carbon burial, because the rate of burial also depends on the residence time of carbon in the mixed

sediment surface layer, and hence carbonate accumulation rates^{23,24}. Increased carbonate preservation, evident at the deeper Site 1237 after 13.9 Myr ago (Fig. 1), may be representative for large areas of the subtropical Pacific. It was probably related to a reduction of the shelf carbonate reservoir due to falling sea level and improved deep-water ventilation, and bears similarity to the rapid deepening of the Calcite Compensation Depth (CCD) near the Eocene–Oligocene boundary when permanent ice sheets appeared on Antarctica²⁵. Indeed, the 1237 Fe and the $\delta^{13}\text{C}$ records at Sites 1146 and 1237 are significantly coherent ($\alpha = 0.1$, not shown) with the 400-kyr eccentricity cycle between 14.3 and 13.1 Myr ago (Fig. 1). Whereas Fe (carbonate dissolution) is in phase with the 400-kyr eccentricity cycle, $\delta^{13}\text{C}$ exhibits a ~50 kyr phase lag (Fig. 1), indicating that the Pacific carbon cycle influenced $\delta^{13}\text{C}$. The $\delta^{13}\text{C}$ minimum at 13.92 Myr ago occurs before the onset of the major $\delta^{18}\text{O}$ increase at 13.91 Myr ago. Interestingly, a minimum in the 400-kyr $\delta^{13}\text{C}$ cycle also precedes the $\delta^{18}\text{O}$ increase at the Oligocene–Miocene boundary²⁶. Thus, the 400-kyr eccentricity cycle appears strongly to influence the carbon cycle²⁷, and may even modulate long-term climate evolution.

Our results indicate that a conjunction of climatic forcing factors ultimately triggered middle-Miocene ice-sheet expansion and global cooling. Between 14.7 and 13.84 Myr ago, ice volume was strongly modulated by obliquity-paced atmospheric moisture transfer. A change in eccentricity cadence (from 400 to 100 kyr) led initially to ice-sheet decay and instability from 14.05 to 13.87 Myr ago. However, the ice-sheet expanded rapidly at 13.87–13.84 Myr ago during an extended interval of low seasonal contrast at low eccentricity that coincided with a prominent $\delta^{13}\text{C}$ increase. We surmise that changing boundary conditions, in particular decreasing p_{CO_2} , primed the dramatic response of the climate system that marked the final step into the 'icehouse' climate. Atmospheric CO_2 fluctuations influencing climate evolution on orbital timescales have not been resolved by previous work^{28,29}, and higher-resolution p_{CO_2} proxy records are urgently needed to clarify the role of atmospheric CO_2 in Cenozoic climate cooling.

METHODS

Chronology. New chronologies were generated by initially matching the 400-kyr and 100-kyr amplitude variations in the $\delta^{18}\text{O}$ series to the latest astronomical solution⁹, and then fine-adjusting individual obliquity-scale cycles. As tuning target, we constructed an eccentricity-tilt composite with no phase shift and equal weight of eccentricity and obliquity (Supplementary Information). Age tiepoints are given in Fig. 1 and Supplementary Tables S3 and S4. We tuned $\delta^{18}\text{O}$ minima to obliquity maxima, since we assumed (1) that relatively warm summers during high obliquity would promote ice-sheet melting in Antarctica, whereas cool summers during low obliquity would favour ice-sheet growth, and (2) that a low summer insolation gradient between low and high latitudes during high obliquity would decrease poleward moisture transport, inhibiting ice-sheet build-up¹⁷. We did not adjust our tuning for possible phase lags between $\delta^{18}\text{O}$ and insolation forcing, since the response time of a smaller middle Miocene Antarctic ice-sheet is unknown. We applied gaussian band-pass filters to the 1146 and 1237 $\delta^{18}\text{O}$ and 1237 Fe series to extract oscillations associated with the 400, 100 and 41 kyr periods. Gaussian band-pass filters were centred at 0.0244 kyr^{-1} (41-kyr period) with 0.005 kyr^{-1} bandwidth (34.0–51.5-kyr period), 0.01 kyr^{-1} (100-kyr period) with 0.003 kyr^{-1} bandwidth (76.9–142.9-kyr period) and 0.0025 kyr^{-1} (400-kyr period) with 0.0005 kyr^{-1} bandwidth (333.3–500.0-kyr period). $\delta^{18}\text{O}$ is phase-locked with obliquity in the 41-kyr band, and the amplitude modulation of the 41-kyr filtered signals closely follows obliquity between 14.7 and 13.5 Myr ago (Fig. 1). The amplitude of the 100-kyr filtered $\delta^{18}\text{O}$ and Fe signals increases after 14.1 Myr ago, exhibiting similar modulation as eccentricity (Figs 1 and 2). Spectral analysis of the $\delta^{18}\text{O}$, Fe and coarse residue time series reveal strong power at all Milankovitch frequency bands, except precession (Supplementary Fig. S4). The spectral peak for short eccentricity shows a characteristic split into 95-kyr and 125-kyr periods. Spectral power in the short and long eccentricity periods are enhanced after tuning to obliquity.

Time-frequency analysis. Temporal changes in amplitude of signal components were estimated over a predefined range of frequencies using a harmonic-filtering algorithm³⁰, which fits sinusoidal waves to a time series by means of least-squares. This method can process unevenly spaced time series directly. To obtain

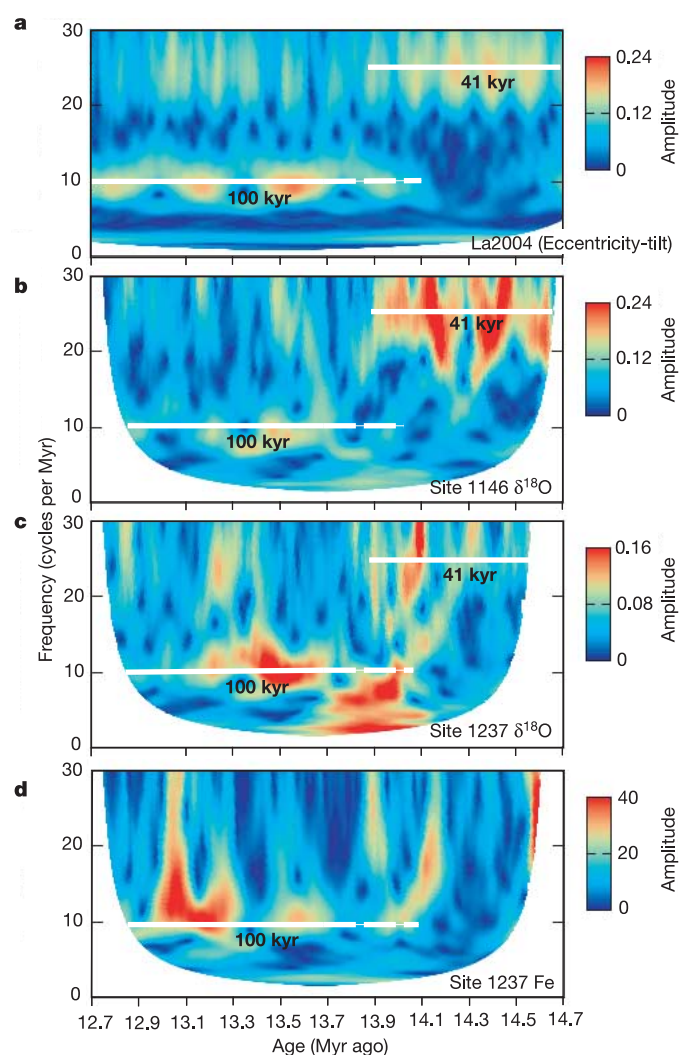


Figure 2 | Signal evolution in time-frequency space of $\delta^{18}\text{O}$ and Fe in ODP Sites 1146 and 1237 and of eccentricity-tilt tuning target. The amplitude evolution in the 100-kyr and 41-kyr bands of $\delta^{18}\text{O}$ follows eccentricity and obliquity. Deep-water cooling and Antarctic ice-sheet expansion coincided with a transition from 41-kyr to 100-kyr periodicity in $\delta^{18}\text{O}$. Signal amplitudes are in same units as data analysed. La2004 refers to the orbital solution in ref. 9 (Laskar *et al.*, 2004).

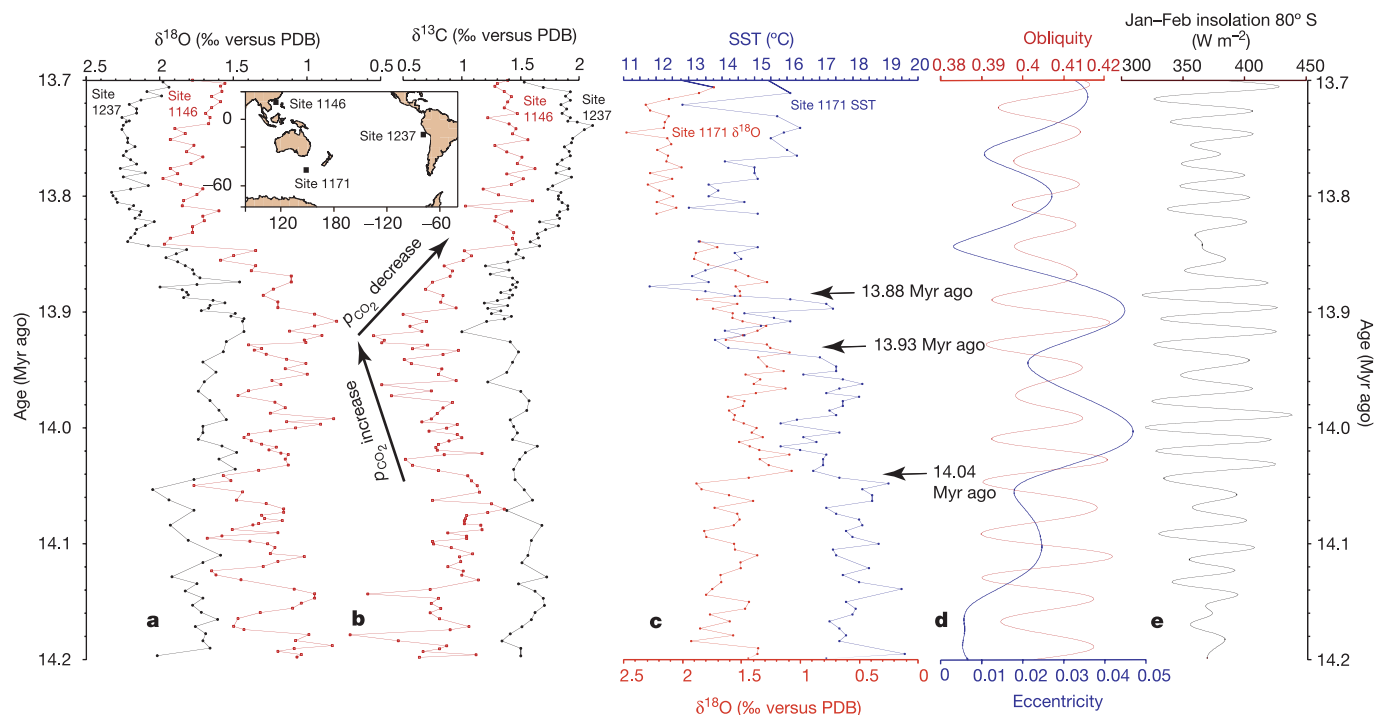


Figure 3 | Expanded view of 14.2–13.7 Myr ago interval. **a, b,** Climate proxies are same as in Fig. 1. A minimum in $\delta^{13}\text{C}$ reflecting a maximum in atmospheric $p\text{CO}_2$ preceded ice-sheet expansion and global cooling. This $\delta^{13}\text{C}$ minimum, indicated by arrows in **b**, marks the turning point of the 400-kyr cycle at 13.92 Myr ago. **c,** Benthic $\delta^{18}\text{O}$ and SSTs at South Tasman

Rise ODP Site 1171 (ref. 21). Arrows indicate prominent, rapid surface cooling events preceding cryosphere expansion. **d,** Obliquity and eccentricity⁹. **e,** January–February insolation at 80° S (ref. 9). The conjunction of low eccentricity and low obliquity resulted in an extended period of low seasonality over Antarctica between 13.87–13.83 Myr ago.

time-dependent amplitude estimates, the input time series is analysed within a moving window of width $T_w = w \times T_f$ where $w = 3$ is a width factor and T_f denotes the signal periodicity of interest (for example, 400 kyr). The resulting amplitude of the best-fit sinusoid is plotted versus the average of the observation times within the current segment. The dependence of window width T_w on frequency leads to a change in temporal resolution with frequency, similar to wavelet analysis (see Supplementary Information for further details).

Isotope analysis. We measured $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ in epifaunal benthic foraminifers (*Planulina wuellerstorfi* or *Cibicides mundulus* except in ten samples from Site 1146, where we analysed *C. barnetti* or *C. incrassatus*). Well-preserved tests were broken into large fragments, cleaned in alcohol in an ultrasonic bath, then dried at 40 °C. Measurements were made with the Finnigan MAT 251 mass spectrometer at the Leibniz Laboratory, Kiel University. The instrument is coupled on-line to a Carbo-Kiel Device (Type I). Samples were reacted by individual acid addition (99% H_3PO_4 at 73 °C). The standard external error is better than $\pm 0.07\text{‰}$ and $\pm 0.05\text{‰}$ for $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$, respectively. Replicate measurements on $\sim 7\%$ of samples indicate mean reproducibility better than $\pm 0.11\text{‰}$ and $\pm 0.13\text{‰}$ for $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$, respectively. Paired measurements in 37 samples indicate no significant offset in $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ between *P. wuellerstorfi* and *C. mundulus*. Results were calibrated using the National Institute of Standards and Technology (Gaithersburg, Maryland) carbonate isotope standard NBS 20 and NBS 19 and 18, and are reported on the Pee Dee belemnite (PDB) scale.

XRF scanning. We performed XRF measurements with 1-cm resolution on the archive halves of the Site 1237 splice using the XRF Core Scanner at the Bremen ODP Core Repository. Overlapping measurements (50–100 cm) were made at correlation points to verify the accuracy and completeness of the splice. We interpreted Fe maxima as intervals of increased carbonate dissolution caused by poor deep-water ventilation and not as intervals of increased terrigenous dust flux or river run-off, based on the covariance of Fe and Mg, and the lack of covariance of Fe and Ti (Supplementary Fig. S5).

Received 21 February; accepted 4 August 2005.

1. Flower, B. P. & Kennett, J. P. Middle Miocene ocean-climate transition: high resolution oxygen and carbon isotopic records from DSDP Site 588A, southwest Pacific. *Paleoceanography* **8**, 811–843 (1993).
2. Flower, B. P. & Kennett, J. P. Middle Miocene deep water paleoceanography in

the southwest Pacific: Relations with East Antarctic ice sheet development. *Paleoceanography* **10**, 1095–1112 (1995).

3. Miller, K. G., Wright, J. D. & Fairbanks, R. G. Unlocking the ice house: Oligocene–Miocene oxygen isotopes, eustasy and margin erosion. *J. Geophys. Res.* **96**, 6829–6848 (1991).
4. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
5. Kennett, J. P. Cenozoic evolution of Antarctic glaciation, the Circum-Antarctic Ocean and their impact on global paleoceanography. *J. Geophys. Res.* **82**, 3843–3860 (1975).
6. Woodruff, F. & Savin, S. Mid-Miocene isotope stratigraphy in the deep sea: high resolution correlations, paleoclimatic cycles, and sediment preservation. *Paleoceanography* **6**, 755–806 (1991).
7. Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
8. Vincent, E. & Berger, W. H. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Broecker, W. S. & Sundquist, E. T.) 455–468 (Geophys. Monogr. Ser. 32, AGU, Washington DC, 1985).
9. Laskar, J. *et al.* A long term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
10. Sugden, D. & Denton, G. Cenozoic landscape evolution of the Convoy Range to Mackay Glacier area, Transantarctic Mountains: Onshore to offshore synthesis. *Geol. Soc. Am. Bull.* **116**, 840–857 (2004).
11. Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperature and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
12. Berger, W. H. & Jansen, E. in *The Polar Oceans and their Role in Shaping the Global Environment* (eds Johannessen, O. M., Muench, R. D. & Overland, J. E.) 295–311 (Geophys. Monogr. Ser. 85, AGU, Washington, DC, 1994).
13. Mudelsee, M. & Schulz, M. The Middle Pleistocene climate transition: onset of 100 kyr cycle lags ice volume buildup by 280 ka. *Earth Planet. Sci. Lett.* **151**, 117–123 (1997).
14. Shackleton, N. J. The 100,000-year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* **289**, 1897–1902 (2000).
15. Imbrie, J. *et al.* On the structure and origin of major glaciation cycles. 2. The 100,000 years cycle. *Paleoceanography* **8**, 699–735 (1993).
16. Raymo, M. E. & Nisancioglu, K. The 41 kyr world: Milankovitch's other unsolved mystery. *Paleoceanography* **18**, 1011 doi:10.1029/2002PA00791 (2003).
17. Loutre, M.-F., Paillard, D., Vimeux, F. & Cortijo, E. Does mean annual insolation have the potential to change the climate? *Earth Planet. Sci. Lett.* **221**, 1–14 (2004).

18. Vimeux, F. *et al.* A 420,000 year deuterium excess record from East Antarctica: Information on past changes in the origin of precipitation at Vostok. *J. Geophys. Res.* **106**, 31863–31873 (2001).
19. Hall, I. R., McCave, N., Shackleton, N. J., Weedon, G. P. & Harris, S. Intensified deep Pacific inflow and ventilation in Pleistocene glacial times. *Nature* **412**, 809–812 (2001).
20. Hall, I. R. *et al.* Paleocurrent reconstruction of the deep Pacific inflow during the middle Miocene: Reflections of Antarctic Ice Sheet growth. *Paleoceanography* **18**, 1040–1051 (2003).
21. Shevenell, A. E., Kennett, J. P. & Lea, D. W. Middle Miocene Southern Ocean cooling and Antarctic cryosphere expansion. *Science* **305**, 1766–1770 (2004).
22. Kump, L. R. & Arthur, M. A. Interpreting carbon-isotope excursions: carbonate and organic matter. *Chem. Geol.* **161**, 181–198 (1999).
23. Müller, P. J. & Suess, E. Productivity, sedimentation rate, and sedimentary organic matter in the oceans. 1. Organic carbon preservation. *Deep-Sea Res. A* **26**, 1347–1362 (1979).
24. Jahnke, R. A. The global ocean flux of particulate organic carbon: Areal distribution and magnitude. *Glob. Biogeochem. Cycles* **10**, 71–88 (1996).
25. Coxhall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).
26. Zachos, J., Shackleton, N. J., Revenaugh, J. S., Pälike, H. & Flower, B. P. Climate response to orbital forcing across the Oligocene-Miocene boundary. *Science* **292**, 274–278 (2001).
27. Wang, P., Tian, J., Cheng, X., Liu, C. & Xu, J. Carbon reservoir changes preceded major ice-sheet expansion at the mid-Brunhes event. *Geology* **31**, 239–242 (2003).
28. Pagani, M., Arthur, M. A. & Freeman, K. H. Miocene evolution of atmospheric carbon dioxide. *Paleoceanography* **14**, 273–292 (1999).
29. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).
30. Ferraz-Mello, S. Estimation of periods from unequally spaced observations. *Astron. J.* **86**, 619–624 (1981).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Shipboard Scientific Parties of the Ocean Drilling Program (ODP) Legs 184 and 202, J. Kennett, U. Röhl, M. Sarthein, A. Shevenell, J. Schönfeld, J. Stoner and J. Zachos for discussions. This research used samples provided by the ODP, and was funded by the Deutsche Forschungsgemeinschaft.

Author Information Data sets are archived at WDC-MARE (<http://www.pangaea.de>). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.E.H. (ah@gpi.uni-kiel.de).

LETTERS

Density of hydrous silicate melt at the conditions of Earth's deep upper mantle

Kyoko N. Matsukage^{1,2}, Zhicheng Jing² & Shun-ichiro Karato²

The chemical evolution of the Earth and the terrestrial planets is largely controlled by the density of silicate melts. If melt density is higher than that of the surrounding solid, incompatible elements dissolved in the melt will be sequestered in the deep mantle^{1,2}. Previous studies on dry (water-free) melts showed that the density of silicate melts can be higher than that of surrounding solids under deep mantle conditions^{3–8}. However, melts formed under deep mantle conditions are also likely to contain some water², which will reduce the melt density. Here we present data constraining the density of hydrous silicate melt at the conditions of ~410 km depth. We show that the water in the silicate melt is more compressible than the other components, and therefore the effect of water in reducing melt density is markedly diminished under high-pressure conditions. Our study indicates that there is a range of conditions under which a (hydrous) melt could be trapped at the 410-km boundary and hence incompatible elements could be sequestered in the deep mantle, although these conditions are sensitive to melt composition as well as the composition of the surrounding mantle.

The density of silicate melts is determined by chemical composition as well as by pressure and temperature. One of the important components that affect melt density is water. However, the influence of water on melt density has been investigated only at low pressures

on silica-rich melts⁹. The effect of pressure on melt density is large, and the major element compositions of melts studied in these previous works are different from those of the melts in the deep mantle, and therefore these previous data cannot be used to infer the density of hydrous melts in the deep upper mantle.

The chemical composition of a melt formed at ~410 km depth (~14 GPa pressure) is likely to be ultramafic^{10–12}. Water content in the melt depends on the water content of the original rock and the degree of partial melting which is controlled by the temperature at which melting occurs. In this study, the major element chemistry of hydrous melt at 410 km depth is estimated by hydrous melting experiments on CaO-MgO-Al₂O₃-SiO₂-pyrolite at pressures of 10–14 GPa (ref. 12) and on CaO-MgO-FeO-Al₂O₃-SiO₂-pyrolite at pressures up to 7.7 GPa (ref. 10). The target melt composition for density measurement in this study is 35.5 wt% SiO₂, 3.3 wt% Al₂O₃, 11.6 wt% FeO, 30.4 wt% MgO and 14.1 wt% CaO. We chose 5 wt% of water (which corresponds to 13–14 mol%).

The floating/sinking sphere technique was used to determine melt density^{3,6}. Single crystals of San Carlos olivine or synthetic diamond were inserted into a Re/Pt double capsule^{13,14} as a density marker, together with a powder of starting materials. The experiments were performed by using a Kawai-type multianvil apparatus with a 14M/8 cell assembly (Figs 1 and 2) at pressures between ~10 GPa and ~16 GPa and at a temperature of 2,170 K (for the pressure and temperature measurements, see Methods).

During the early stages of this research, we recognized that olivine tends to react with a hydrous ultramafic melt; the olivine disappeared in most cases. Consequently, in the subsequent studies we used a

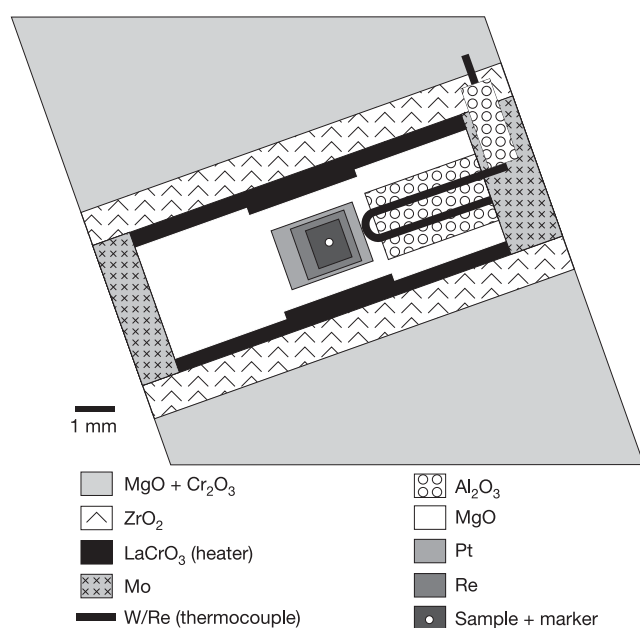


Figure 1 | A schematic cross-section of the cell assembly.

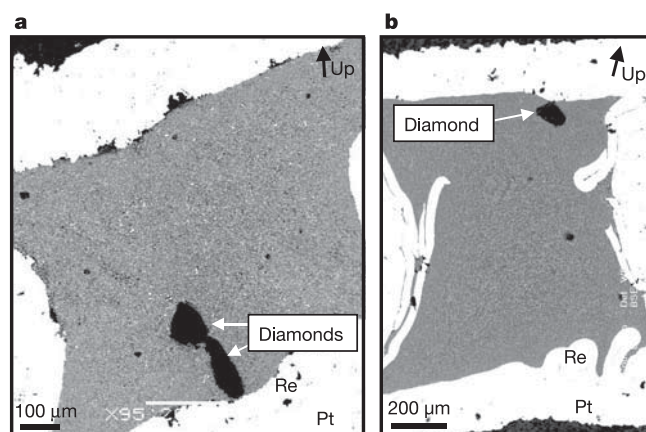


Figure 2 | Back-scattered electron images of run products. **a**, Run K258, 11.8 GPa; diamond density markers have sunk. **b**, Run K248, 16.2 GPa; diamond density marker has floated.

¹Department of Environmental Sciences, Ibaraki University, Mito, Ibaraki 310-0056, Japan. ²Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520, USA.

Table 1 | Chemical compositions of starting materials, run conditions and the result

	Run no.	Pressure (GPa)	Duration (s)	Marker density (g cm ⁻³)	Result
s3-a*	K274	16.2	60	3.557	Float
	K282	14.6	60	3.544	Sink
	K285	13.3	60	3.534	Sink
	K287	15.5	60	3.552	Neutral
s7-a†	K268	13.3	40	3.534	Sink
	K273	14.0	40	3.539	Neutral
	K283	14.6	60	3.544	Neutral
	K286	15.2	60	3.549	Float
s5-a‡	K255	16.2	60	3.557	Float
	K257	13.3	20	3.534	Neutral
	K258	11.8	20	3.521	Sink
s6-a§	K260	13.3	20	3.534	Float
	K263	11.8	20	3.521	Neutral
	K264	10.9	20	3.514	Neutral
	K267	10.0	20	3.506	Sink

Single crystals of diamond 80–100 μm across were used as density markers. Densities of diamond at high pressure and high temperature were calculated using the third-order Birch–Murnaghan equation of state using the data listed in ref. 5. Starting materials are shown bold, with their compositions listed below.
* SiO₂ = 35.1 wt%, Al₂O₃ = 3.2 wt%, FeO = 15.2 wt%, MgO = 28.5 wt%, CaO = 13.1 wt%, H₂O = 5 wt%.
† SiO₂ = 34.3 wt%, Al₂O₃ = 3.2 wt%, FeO = 19.9 wt%, MgO = 25.7 wt%, CaO = 11.9 wt%, H₂O = 5 wt%.
‡ SiO₂ = 33.9 wt%, Al₂O₃ = 1.6 wt%, FeO = 24.9 wt%, MgO = 23.6 wt%, CaO = 11.1 wt%, H₂O = 5 wt%.
§ SiO₂ = 33.5 wt%, Al₂O₃ = 3.1 wt%, FeO = 25.5 wt%, MgO = 22.5 wt%, CaO = 10.5 wt%, H₂O = 5 wt%.

diamond as a density marker. Diamond is, however, considerably denser than olivine and melts with Earth-like bulk chemical composition, such as the target melt, under the conditions of 410 km depth. Therefore, we have chosen to study the density of hydrous ultramafic melts with various FeO contents. When some FeO is added to the system, the density of the melt increases, so that it becomes comparable to that of a diamond. By changing the FeO content systematically, one can determine the density of the melt as a function of FeO content. Using the relation between melt density and FeO content, we determine the melt density with the FeO content appropriate for the target melt with FeO = ~11 wt% (0.08 mole fraction); we call this ‘mantle melt’ in the following. The results are summarized in Table 1 and Fig. 3. We found the density crossover between the four Fe-rich hydrous melts and diamond at various pressures. The densities and pressures of density crossover between diamond and melts are 3.55 g cm⁻³ at 15.5 GPa in melt s3-a, 3.54 g cm⁻³ at 14.3 GPa in melt s7-a, 3.53 g cm⁻³ at 13.3 GPa in melt s5-a, and 3.52 g cm⁻³ at 11.3 GPa in melt s6-a, respectively.

To determine the density of mantle melt at the pressure condition of 410 km depth (14 GPa) from our data at various pressures, we first made corrections for pressure effects and calculated the densities of the four Fe-rich melts at 14 GPa and 2,170 K by using the third-order Birch–Murnaghan equation of state (see Methods). Then, we calculated the density of hydrous melt with the mantle value of FeO content using the relation between four Fe-rich melt densities at 14 GPa and 2,170 K and their FeO contents (Fig. 4). The relation between the (hydrous) melt density and FeO content was determined by assuming linear additivity of partial molar volumes¹⁵, $\rho = (\sum x_i M_i) / (\sum x_i V_i) = (\sum x_i M_i) / [A(1 - x_{\text{FeO}}) + Bx_{\text{FeO}}]$, where x_i and M_i are mole fraction and molecular weight of the i th component in the melt, respectively (we consider six components; SiO₂, Al₂O₃, FeO, MgO, CaO and H₂O), and A and B are relevant parameters that depend on the molar volumes. The parameters A and B were determined by weighted nonlinear least-squares fit. The density of the mantle melt with 5 wt% of water at 14 GPa and 2,170 K is determined to be ~3.42 g cm⁻³ (Fig. 4).

To estimate the effect of water on melt density, the density of our

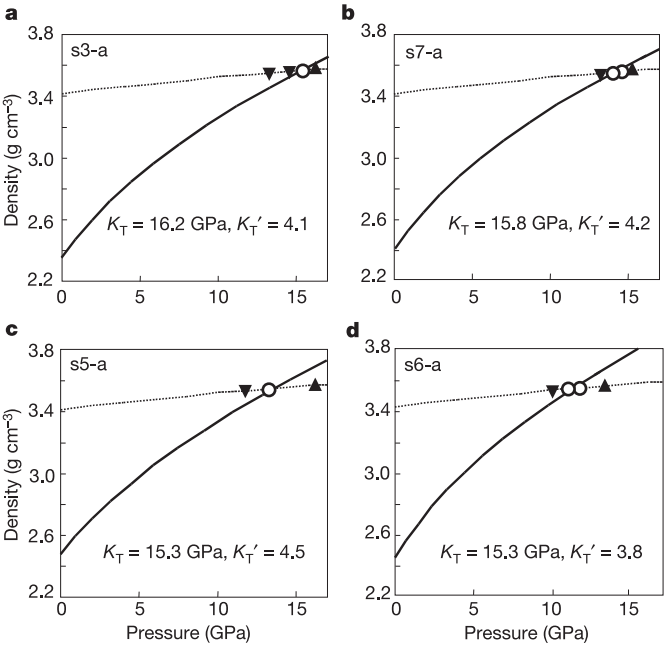


Figure 3 | Compression curves of the hydrous melts and density marker (diamond) at 2,170 K. The compression curve of diamond⁵ (dotted lines) and the hydrous melts (solid lines) was calculated by a third-order Birch–Murnaghan equation of state. The shape of a data point indicates the action of the diamond density marker: down triangle, sinks; up triangle, floats; circle, neutral.

hydrous melt is compared with that of dry (water-free) melt (Fig. 4). The density data for dry melts are from the previous studies^{5–7} on picritic melt, MA, IT8720, MORB and PHN1611. Our hydrous melts are less dense than dry melts with the same FeO content, but the difference is relatively small. We used the relation $\rho = (\sum x_i M_i) / (\sum x_i V_i)$ to infer the partial molar volume of water at 2,170 K and 14 GPa, where V_i is the partial molar volume of components. The partial molar volume of water was estimated to be $\sim 8 \pm 2 \text{ cm}^3 \text{ mol}^{-1}$ at ~14 GPa (density of $\sim 2.5 \pm 0.5 \text{ cm}^3 \text{ mol}^{-1}$). This value of partial molar volume is notably lower than those determined previously at lower pressures. In the previous study, the partial molar volume of water was estimated to be $\sim 27.8 \text{ cm}^3 \text{ mol}^{-1}$ in albite melt at 1,673 K and room pressure⁹, and $\sim 21 \text{ cm}^3 \text{ mol}^{-1}$ in basaltic melt at 1,223 K and 0.77 GPa (ref. 16). We compare the compressibility of water with that of other components (SiO₂, Al₂O₃, FeO, MgO, CaO). If we define $\xi_i \equiv V_{i(1\text{atm}, 2,170\text{K})} / V_{i(14\text{GPa}, 2,170\text{K})}$ (where $V_{i(P,T)}$ is the partial molar volume of the i th species at pressure, P , and temperature, T), we find that $\xi_{\text{H}_2\text{O}} \approx 3.7$ whereas $\xi_{\text{SiO}_2} \approx 1.5$ and $\xi_{\text{MgO}} \approx 1.1$. This indicates that H₂O is more compressible than the other components. These calculations suggest that the degree to which the addition of water reduces density is significantly lowered at high pressures, although water still reduces the total melt density at high pressures.

Uncertainties in the temperature of the real Earth’s interior and in the water content of partial melt influence the estimated melt density a great deal. The actual temperature at which partial melting occurs at ~410 km depth in the Earth² is considered to be lower than our experimental condition; it is estimated to be probably ~1,800 K (ref. 17). Owing to a large difference in thermal expansion between melts and solid silicates^{5–7}, this also changes the relative density between melts and solids. The densities of melt and solids corrected for 1,800 K are also shown in Fig. 4. Our calculation suggests that the density of hydrous melt containing 5 wt% water at ~410 km is close to that of the upper mantle but less than that of the transition zone at 1,800 K, but the density is less than that of the upper mantle at 2,170 K.

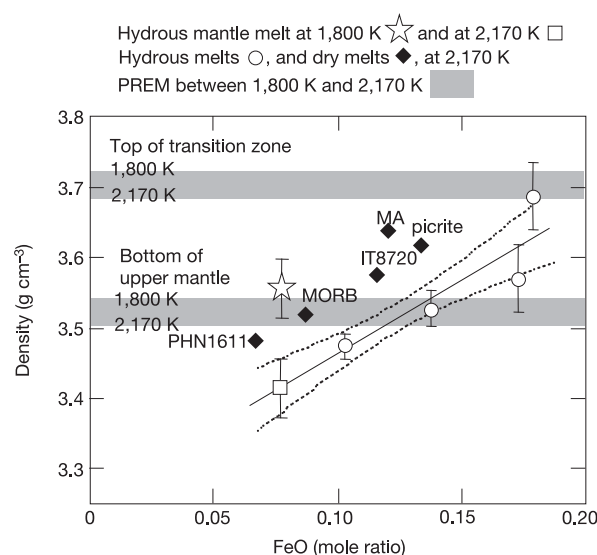


Figure 4 | Densities of various hydrous melts with 5 wt% H₂O and dry melts at 410 km depth as a function of the FeO content. Calculation of the errors for the hydrous melt densities is described in Methods (see (1) and (3) of 'Error estimations of melt density'). The error bars for mantle melts and the envelope shown by broken lines indicate one standard deviation. The densities for the dry mafic and ultramafic melts, that is MORB, PHN1611, IT8720, MA and picrite, at 2,170 K and at 14 GPa are calculated from the data of previous studies at higher temperature between 2,253 and 2,773 K (refs 5–7) after correction for the effects of thermal expansion. The densities of upper mantle and the transition zone near 410-km depth are from PREM¹⁸ with the corrections for thermal expansion corresponding to different geotherms.

The density of hydrous mantle melt is compared with the densities of surrounding silicate solids in Fig. 4. The density of materials in the upper mantle and transition zone near 410 km were from PREM¹⁸, and the influence of temperature was included assuming an average temperature of 1,800 K at 410 km using the experimentally determined thermal expansion¹⁷. The density of pyrolite ($\rho = 3.54 \text{ g cm}^{-3}$ at 2,170 K and 3.58 g cm^{-3} at 1,800 K; ref. 19) is slightly higher than that for PREM²⁰ at the bottom of upper mantle. However, modestly depleted peridotites have densities similar to PREM because of lower Fe and garnet contents.

The density of a silicate melt will be higher than that of the surrounding solid if water content is small. However, the density contrast between a solid and a melt depends also on temperature, because the thermal expansion of a silicate melt is much larger than that of a silicate solid²¹. Consequently, there is a trade-off between water content and temperature when calculating the conditions for density crossover: the critical water content for the density crossover depends on temperature. This point is shown in Fig. 5, where the conditions for density crossover are shown for a range of partial molar volumes of water. For a given temperature, there is an upper limit for water content for which the density crossover occurs. The critical water content is $\sim 6 \text{ wt\%}$ for $T = 1,800 \text{ K}$ and $V_{\text{H}_2\text{O}} = 8 \text{ cm}^3 \text{ mol}^{-1}$. Note, however, that the uncertainties in the current estimate of the partial molar volume of water are large: the density crossover could occur at a water content of $\sim 20 \text{ wt\%}$ and $\sim 2 \text{ wt\%}$ if the partial molar volume of water is $6 \text{ cm}^3 \text{ mol}^{-1}$ and $10 \text{ cm}^3 \text{ mol}^{-1}$, respectively.

In addition to the uncertainty in the experimental determination of the partial molar volume of water, we should also point out several important issues that need to be addressed to make a stronger conclusion about the density crossover. (1) The major element chemistry of the melt used in this study is silica-poor and Ca-rich compared to peridotites or komatiites¹². Melts formed at different depths, or melts formed at a higher degree of melting, could have

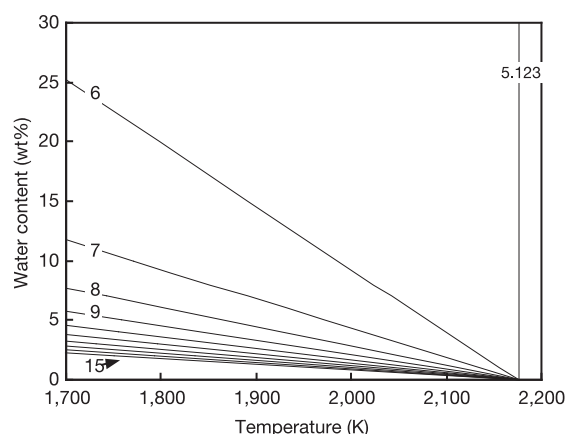


Figure 5 | The condition of density crossover between the hydrous melt and surrounding silicate solid at 410 km depth for a range of the partial molar volume of water. Numbers on each line indicate a value of the partial molar volume of water (in $\text{cm}^3 \text{ mol}^{-1}$). For a given partial molar volume and temperature, the density crossover will occur if the water content is less than a critical value. The temperature at 410 km is estimated to be $\sim 1,800 \text{ K}$ (ref. 17). For this temperature, the density crossover will occur if the water content is less than $\sim 6 \text{ wt\%}$ for a partial molar volume of $8 \text{ cm}^3 \text{ mol}^{-1}$. However, if the uncertainty in the partial molar volume is taken into account ($8 \pm 2 \text{ cm}^3 \text{ mol}^{-1}$), then for $T = 1,800 \text{ K}$, the critical water content will be $6^{+14}_{-4} \text{ wt\%}$.

more peridotitic composition and hence have different densities²². (2) The influence of oxygen fugacity may also be important. In our present study, we used a rhenium capsule that defines a relatively high oxygen fugacity^{14,23}. At lower oxygen fugacities, the oxidation state of iron is likely to be different, and that will make such melts denser than those in the present study. (3) The reference density of solid materials just above the 410-km discontinuity needs to be better constrained (density depends critically on the degree of depletion of these materials). (4) The thermal expansion of hydrous melt under these deep mantle conditions is not well constrained. Low-pressure data on the influence of water on thermal expansion show that thermal expansion of a melt increases significantly with water content⁹. The effects of water on the thermal expansion of melt need to be determined at high pressures. Further studies are needed on these issues to address the relative densities of hydrous melt and surrounding minerals near the 410-km boundary.

METHODS

Materials. The starting materials were prepared from reagent grade SiO_2 , Al_2O_3 , FeO , MgO , $\text{Mg}(\text{OH})_2$ and $\text{Ca}(\text{OH})_2$, which were mixed in an appropriate ratio and hand-ground using an agate mortar and pestle. Round single crystals of diamond, 80–100 μm in diameter, and olivine, 100–150 μm in diameter, were used as density markers. After an experiment, the run products were sectioned parallel to the gravity direction during the experiment, and polished for observation and analysis. Observations with an optical microscope and a scanning electron microscope (SEM) in the back-scattered electron imaging mode showed that the run products were mixtures of quench crystals, glass and a small amount of bubbles, and were free of chemical and textural heterogeneity (Fig. 2). These observations confirm the total melting of starting materials and the existence of water during heating.

Pressure and temperature measurements. Pressure and temperature for each experiment were estimated through the load–pressure calibration and through a thermocouple reading. The pressure–load calibration was made using the phase transformation of SiO_2 from coesite to stishovite and Mg_2SiO_4 from olivine to wadsleyite, based on the results of *in situ* measurements^{24,25}. The error in the pressure estimate comes from uncertainties in the absolute pressure scale and the imperfect reproducibility in the experiments, and is estimated to be about $\pm 0.5 \text{ GPa}$. The uncertainties in the temperature estimate come from the uncertainties in the pressure effects on the $\text{W}_{50}\text{Re}–\text{W}_{26}\text{Re}$ thermocouple e.m.f.²⁶, and are estimated to be about $\pm 100 \text{ K}$.

Density correction and equation of state. Densities of melt determined at

various pressures were corrected to a single pressure to determine the influence of Fe content on density. This was made using the third-order Birch–Murnaghan equation of state

$$P = 3/2 K_{T0}[(\rho/\rho_0)^{2/3} - 1](\rho/\rho_0)^{5/3} \{1 + 3/4 (K_{T0}' - 4)[(\rho/\rho_0)^{2/3} - 1]\} \quad (1)$$

where K_{T0} , K_{T0}' , ρ_0 and ρ are the isothermal bulk modulus at room pressure, the pressure derivative of K_{T0} at room pressure, the room-pressure density and the high-pressure density, respectively (all quantities are at the same temperature). ρ_0 and K_{T0} were calculated from the partial molar volumes of each component and their pressure derivatives assuming the linear additivity of partial molar volumes¹⁵. We used the data from refs 9 and 27–29 to calculate ρ_0 and K_{T0} , and used equation (1) to determine K_{T0}' . The densities at different pressures at the same temperature were calculated using equation (1). The effects of temperature were corrected using the thermal expansivities or equivalently using the Mie–Grüneisen equation of state. The data for thermal expansivities are from refs 6 and 9.

Error estimates of melt density. There are several sources of uncertainties in the estimates of hydrous melt density.

(1) When we determine the density of the melt by bracketing the density, there is a finite density interval that introduces a source of error. The density of melt can also be inferred from the ‘neutral’ results, in which the density standard did not move more than ~ 1 mm in a capsule. In these cases, the density differences between the marker material and the melt must be less than a certain value that can be estimated from Stokes’ law. A smaller value of these two different sources of uncertainties was chosen for the error for the melt density. The typical error caused by this process is approximately $\pm 1\%$ (see Fig. 4).

(2) The error due to the extrapolation or interpolation of the data using the equation of state. One source of error is the error in the room-pressure partial molar volume of H_2O , which is not well constrained^{9,30}. However, because the melt densities at 14 GPa are constrained by the actual experimental data, the error in the room-pressure molar volume of H_2O does not result in a large error in the 14 GPa melt density (less than $\pm 0.3\%$), although it affects the value of K' and hence the melt density at largely different pressures will be affected significantly.

(3) The error associated with the extrapolation of high FeO content data to a lower FeO content. This is due to the error in parameters A and B . This causes an error of about $\pm 1.3\%$ (Fig. 4).

(4) The uncertainties in the use of the thermal expansion of dry melt in calculating the melt density at different temperatures (Fig. 4). The thermal expansion of dry melt was used for hydrous melts, because there are no bulk thermal expansion data for hydrous melts. According to ref. 9, thermal expansion increases systematically with water content (addition of ~ 5 wt% of H_2O results in an increase of thermal expansion by $\sim 25\%$ at 0.5–0.8 GPa). However, thermal expansion should be much less at higher pressures (compare results of refs 6 and 9), so the effect of water is probably less. But there are no data on thermal expansion on hydrous silicate melts at high pressures (~ 14 GPa).

(5) Uncertainty in the partial molar volume of H_2O also causes uncertainties in the critical water content for density crossover. The uncertainty in partial molar volume is about $\pm 20\%$. The influence of uncertainties on the conditions for density crossover can be seen in Fig. 5.

Received 3 March; accepted 19 September 2005.

1. Stolper, E. M., Walker, D., Hager, B. H. & Hays, J. F. Melt segregation from partially molten source regions: the importance of melt density and source region size. *J. Geophys. Res.* **86**, 6261–6271 (1981).
2. Bercovici, D. & Karato, S. Whole-mantle convection and the transition-zone water filter. *Nature* **425**, 39–44 (2003).
3. Agee, C. B. & Walker, D. Olivine flotation in mantle melt. *Phys. Earth Planet. Inter.* **114**, 315–324 (1988).
4. Ohtani, E., Suzuki, A. & Kato, T. Flotation of olivine in the peridotite melt at high pressure. *Proc. Jpn Acad.* **69**, 23–28 (1993).
5. Suzuki, A., Ohtani, E. & Kato, T. Flotation of diamond in mantle melt at high pressure. *Science* **269**, 216–218 (1995).

6. Suzuki, A., Ohtani, E. & Kato, T. Density and thermal expansion of a peridotite melt at high pressure. *Phys. Earth Planet. Inter.* **107**, 53–61 (1998).
7. Suzuki, A. & Ohtani, E. Density of peridotite melts at high pressure. *Phys. Chem. Miner.* **30**, 449–456 (2003).
8. Ohtani, E. & Maeda, M. Density of basaltic melt at high pressure and stability of the melt at the base of the lower mantle. *Phys. Earth Planet. Inter.* **193**, 69–75 (2001).
9. Ochs, F. A. III & Lange, R. A. The partial molar volume, thermal expansivity, and compressibility of H_2O in $NaAlSi_3O_8$ liquid: new measurements and an internally consistent model. *Contrib. Mineral. Petrol.* **129**, 155–165 (1997).
10. Inoue, T. & Sawamoto, H. in *High-pressure Research: Application to Earth and Planetary Sciences* (eds Shono, Y. & Manghnani, H.) 323–331 (Terra Scientific, Tokyo/American Geophysical Union, Washington DC, 1992).
11. Inoue, T. Effect of water on melting phase relations and melt composition in the system Mg_2SiO_4 – $MgSiO_3$ – H_2O up to 15 GPa. *Phys. Earth Planet. Inter.* **85**, 237–263 (1994).
12. Litasov, K. & Ohtani, E. Phase relations and melt compositions in CMAS–pyrolite– H_2O system up to 25 GPa. *Phys. Earth Planet. Inter.* **134**, 105–127 (2002).
13. Takahashi, E., Shimazaki, T., Tsuzaki, Y. & Yoshida, H. Melting study of a peridotite KLB-1 to 6.5 GPa and the origin of basaltic magmas. *Phil. Trans. R. Soc. Lond. A* **342**, 105–120 (1993).
14. Matsukage, K. N. & Kubo, K. Chromian spinel during melting experiments of dry peridotite (KLB-1) at 1.0–2.5 GPa. *Am. Mineral.* **88**, 1271–1278 (2003).
15. Bottinga, Y. & Weill, D. F. Densities of liquid silicate systems calculated from partial molar volumes of oxide components. *Am. J. Sci.* **269**, 169–182 (1970).
16. Burnham, C. W. & Davis, N. F. The role of H_2O in silicate melts: I P–V–T relations in the system $NaAlSi_3O_8$ – H_2O to 10 kilobars and 1000°C. *Am. J. Sci.* **270**, 54–79 (1971).
17. Ito, E. & Katsura, T. A temperature profile of the mantle transition zone. *Geophys. Res. Lett.* **16**, 425–528 (1989).
18. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
19. Matsukage, K. N., Nishihara, Y. & Karato, S. Seismological signature of chemical differentiation of Earth’s upper mantle. *J. Geophys. Res.* (in the press).
20. Nishihara, Y. & Takahashi, E. Phase relation and physical properties of an Al-depleted komatiite to 23 GPa. *Earth Planet. Sci. Lett.* **190**, 65–77 (2001).
21. Fei, Y. in *Mineral Physics & Crystallography, A Handbook of Physical Constants* (ed. Ahrens, T. J.) 29–44 (American Geophysical Union, Washington DC, 1995).
22. Agee, C. Crystal-liquid density inversions in terrestrial and lunar magmas. *Phys. Earth Planet. Inter.* **107**, 63–74 (1998).
23. Nishihara, Y., Shinmei, T. & Karato, S. Grain-growth kinetics in wadsleyite: effects of chemical environment. *Phys. Earth Planet. Inter.* (in the press).
24. Zhang, J., Li, B., Utsumi, W. & Liberman, R. C. In situ X-ray observations of the coesite–stishovite transition: reversed phase boundary and kinetics. *Phys. Chem. Miner.* **23**, 1–10 (1996).
25. Morishima, H. *et al.* The phase boundary between α - Mg_2SiO_4 and β - Mg_2SiO_4 determined by in-situ X-ray observation. *Science* **265**, 1202–1203 (2004).
26. Li, J., Hadidiacos, C., Mao, H.-K., Fei, Y. & Hemley, R. J. Effect of pressure on thermocouples in a multi-anvil apparatus. *High Press. Res.* **23**, 389–401 (2003).
27. Lange, R. A. & Carmichael, I. S. E. Density of Na_2O – K_2O – CaO – MgO – FeO – Fe_2O_3 – Al_2O_3 – TiO_2 – SiO_2 liquids: new measurements and derived partial molar properties. *Geochim. Cosmochim. Acta* **51**, 2931–2946 (1987).
28. Lange, R. A. A revised model for the density and thermal expansivity of K_2O – Na_2O – CaO – MgO – Al_2O_3 – SiO_2 liquids from 700 to 1900 K: extension to crustal magmatic temperatures. *Contrib. Mineral. Petrol.* **130**, 1–11 (1997).
29. Ochs, F. A. III & Lange, R. A. The density of hydrous magmatic liquid. *Science* **283**, 1314–1317 (1999).
30. Richet, P. *et al.* Water and the density of silicate glasses. *Contrib. Mineral. Petrol.* **138**, 337–347 (2000).

Acknowledgements We thank Y. Nishihara for advice on experimental techniques and discussions, and H. Sumiya for the supply of single crystals of diamond. This work was supported by NSF (S.K.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.N.M. (kmatsu@mx.ibaraki.ac.jp).

LETTERS

The entomological inoculation rate and *Plasmodium falciparum* infection in African children

D. L. Smith¹, J. Dushoff^{1,2}, R. W. Snow^{3,4} & S. I. Hay^{3,5}

Malaria is an important cause of global morbidity and mortality. The fact that some people are bitten more often than others has a large effect on the relationship between risk factors and prevalence of vector-borne diseases^{1–3}. Here we develop a mathematical framework that allows us to estimate the heterogeneity of infection rates from the relationship between rates of infectious bites and community prevalence. We apply this framework to a large, published data set that combines malaria measurements from more than 90 communities⁴. We find strong evidence that heterogeneous biting or heterogeneous susceptibility to infection are important and pervasive factors determining the prevalence of infection: 20% of people receive 80% of all infections. We also find that individual infections last about six months on average, per infectious bite, and children who clear infections are not immune to new infections. The results have important implications for public health interventions: the success of malaria control will depend heavily on whether efforts are targeted at those who are most at risk of infection.

Economic, public-health, and medical advances over the past 100 years halved the global geographical extent of malaria⁵. Nonetheless, population growth and failure to prevent infection or manage disease in this reduced extent means that *Plasmodium falciparum* remains a leading cause of global morbidity⁶ and mortality^{4,7,8}. Central to the future of malaria control is more effective understanding of the relationships between risk factors, the frequency and longevity of infection, and disease outcome⁹. These relationships have been debated extensively in relation to disease control^{10,11}.

An important aspect of this debate is the quantitative relationship that inevitably exists between the proportion of people who are infected with *P. falciparum*, (the parasite ratio, PR) and the rate at which people are bitten by infectious mosquitoes (the entomological inoculation rate, EIR). The EIR is rarely recorded in Africa⁴, limiting informed debate on the relationship between vector biology, transmission intensity, clinical disease and mortality risks. Conversely, the PR is a widely measured index of infection risk, allowing a more detailed investigation of the interaction between transmission intensity, age, and disease burden. Using mathematical models, we seek here to explore the factors that determine the complex, nonlinear association between EIR and PR, and use these models to highlight key features of the biology of malaria transmission that relate to future disease control.

Ross first developed a mathematical model for the relationship between EIR and PR¹², but his model, as extended by Macdonald, performed poorly in the African Savannah¹³. Since then, several modifications to Ross' original model have been proposed that are now considered important in malaria epidemiology. Informed by a century of theory, we have developed a large set of mathematical

models, fitted them to a comprehensive data set, and selected the best of these models to describe the relationship between EIR and PR (Methods).

It was assumed by Ross that human populations were homogeneous, but in fact some people are bitten by mosquitoes more than others because of proximity to larval habitat¹⁴, differential attractiveness to mosquitoes¹⁵, or other reasons. Moreover, some people are more susceptible to infection, per bite. Heterogeneous infection rates have important implications for the dynamics and control of malaria^{1–3}, and heterogeneity fundamentally changes the relationship between EIR and PR¹⁴. Those who are infected most play a role in malaria transmission that is analogous to the role of the most sexually active in transmission of sexually transmitted diseases¹⁶. We assume that relative infection rates follow a Γ distribution, with mean 1 and variance $1/k$.

Ross assumed that infections clear at a constant rate, regardless of subsequent infections. The assumption was challenged by a growing consensus that superinfection with *P. falciparum* would increase the time to clear an infection^{17–20}. Let $\mathcal{E}$ denote the annual EIR, b the transmission efficiency (the probability that a bite by an infectious mosquito results in an infection—either a new infection or superinfection) and $1/r$ the expected time to clear each infection. Assuming superinfection and assuming infections clear independently, clearance occurs at the rate $b\mathcal{E}/(e^{b\mathcal{E}/r} - 1)$ (refs 17–19); thus, the time to clearance is longer when annual EIR is higher. We also considered immunity to reinfection by comparing SIS to SIRS dynamics (Methods). The full set of candidate models included SIS and SIRS models combined with heterogeneous infection and superinfection.

We identified 119 empirical estimates of EIR matched to coincidental measures of PR in African children under 15 years of age⁴; these data should be regarded with circumspection, as they were collected using different methods and for other purposes. One hundred and nine pairs measured positive EIR and PR. Ninety-one of these pairs also reported the sample sizes for the PR estimate. Two other modifications to these models were considered. First, microscopy errors bias estimates of PR and affect the analysis²¹. As the sensitivity and specificity of microscopy is not known, they were fitted along with the other model parameters. Second, the age ranges of the children sampled differed among studies. This introduces a potential bias if estimated prevalence varied substantially with age. We fitted each model with and without age corrections and microscopy errors (Methods).

Each candidate model was fitted to the data by maximum likelihood using R (ref. 22) and compared to a log-linear model^{5,23}. Based on these fits, the best overall model was selected using Akaike's information criterion (AIC)²⁴ (Table 1, Fig 1a). The best overall

¹Fogarty International Center, National Institutes of Health, Building 16, 16 Center Drive, Bethesda, Maryland 20892, USA. ²Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA. ³Malaria Public Health & Epidemiology Group, Centre for Geographic Medicine, KEMRI (in the grounds of the Kenyatta Hospital), PO Box 43640, 00100 Nairobi GPO, Kenya. ⁴Centre for Tropical Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK. ⁵TALA Research Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Table 1 | The models and fitted parameters

PR	ΔAIC	b/r	$1/k$	$b\phi$	α	β	a	b	Name
$1 - \left(1 + \frac{b\mathcal{E}}{rk}\right)^{-k}$	0	0.45	4.2		96%	12%	0.008	-0.0016	$\int \text{SI}^{\circ}\text{S}$
$\int_0^{\infty} \left(\frac{1 - e^{-s\mathcal{E}/r}}{1 + b\phi\mathcal{E}e^{-s\mathcal{E}/r}} \right) g(s)ds$	2	0.45	4.2	0	96%	12%	0.008	-0.016	$\int \text{SI}^{\circ}\text{RS}$
$\frac{b\mathcal{E}}{b\mathcal{E} + r}$	213	0.16			78%	16%	0.01	-0.026	SIS
$\frac{1 - e^{-b\mathcal{E}/r}}{1 + b\phi\mathcal{E}e^{-b\mathcal{E}/r}}$	557	0.1		0.05	76%	21%	0.012	-0.028	$\text{SI}^{\circ}\text{RS}$
$1 - e^{-b\mathcal{E}/r}$	612	0.077			75%	21%	0.01	-0.03	$\text{SI}^{\circ}\text{S}$
		c	d						
$c\log\mathcal{E} + d$	185	0.084	0.32				0.008	-0.012	Log-linear

Models are ranked by AIC. The analysis found strong evidential support for the model $\int \text{SI}^{\circ}\text{S}$ described by equation (1), a model with heterogeneous infection rates (the symbol $\int$ denotes heterogeneous infection) and superinfection (the default assumption is constant clearance; $^{\circ}$ implies superinfection). The models that lacked heterogeneous infection, SIS, $\text{SI}^{\circ}\text{S}$, $\text{SI}^{\circ}\text{RS}$, and the log-linear model fit poorly, as judged by AIC. (We report the ΔAIC values.) Moreover, the parameters for the model $\int \text{SI}^{\circ}\text{S}$ are biologically meaningful, unlike those for the log-linear model. Finally, the model $\int \text{SI}^{\circ}\text{RS}$ fitted no better than $\int \text{SI}^{\circ}\text{S}$, and the fitted duration of immunity to infection was zero.

model was a simple function that incorporated heterogeneous infection rates, with no immunity to reinfection:

$$\text{PR} = 1 - \left(1 + \frac{b\mathcal{E}}{rk}\right)^{-k} \quad (1)$$

Equation (1) fitted better than the log-linear model (another two-parameter model) and had very strong evidential support over the log-linear model, with ΔAIC of 185 (refs 5, 23). Moreover, equation (1) was selected over the log-linear model in every case when the analysis was repeated on 1,000 bootstrapped data sets.

The model has six fitted parameters, including sensitivity, specificity, age-corrections, k and b/r : b and r are exactly co-linear—the mathematical relationship between EIR and PR in these models depends only on their ratio. For the best overall model, the fitted sensitivity and specificity were 95.8% and 88.4%, respectively. The estimates were different for each model. A sensitivity analysis demonstrates that PR is much more sensitive to changes in k , which determines variance in infection rates, when annual EIR is greater than about 1, but PR is more sensitive to b/r when EIR is lower (Fig. 1b).

A simple summary of heterogeneous infection is the fraction of all infections received by the subpopulation that is infected most frequently; for $1/k = 4.2$, 20% of the population receives 80% of all infections, similar to a single study from Tanzania in which 20% of the population received 80% of all bites³. This represents an average across the 91 populations sampled—the distribution of infection rates in a particular population may be more or less heterogeneous, depending on the local ecology¹⁴.

The fitted parameter b/r is the product of transmission efficiency and persistence times; alternatively, it is the expected duration of an infection, per infectious bite. If transmission efficiency were perfect, the best-fit parameter would correspond to a duration of infection of approximately 166 days; if transmission efficiency were approximately 50%, then persistence would be 11 months. These estimates are consistent with estimates of persistence from simple infections induced for malaria therapy²⁵ and with recent studies of persistence for natural infections^{26,27}.

Consistent with other studies^{28,29} and the notion that immunity to clinical illness develops after repeated infection in early childhood, we found no evidence for immunity to infection among these populations of African children, as reflected in the relationship between EIR and PR. A direct comparison of SIS and SIRS models (Methods; equation (4) versus equations (6) or (7)) demonstrated that mathematical models for malaria infection in children should be SIS and not SIRS because children do not become immune to

infection after clearing a single infection, but immunity requires repeated infection or possibly some change in immune function with age.

On the other hand, the analysis did reveal a strong decline in prevalence associated with the maximum age of the sample population. The lower bound for age was associated with a 0.8% increase in prevalence for each year of age, and the upper bound was associated with a 1.6% decrease in prevalence. On closer scrutiny,

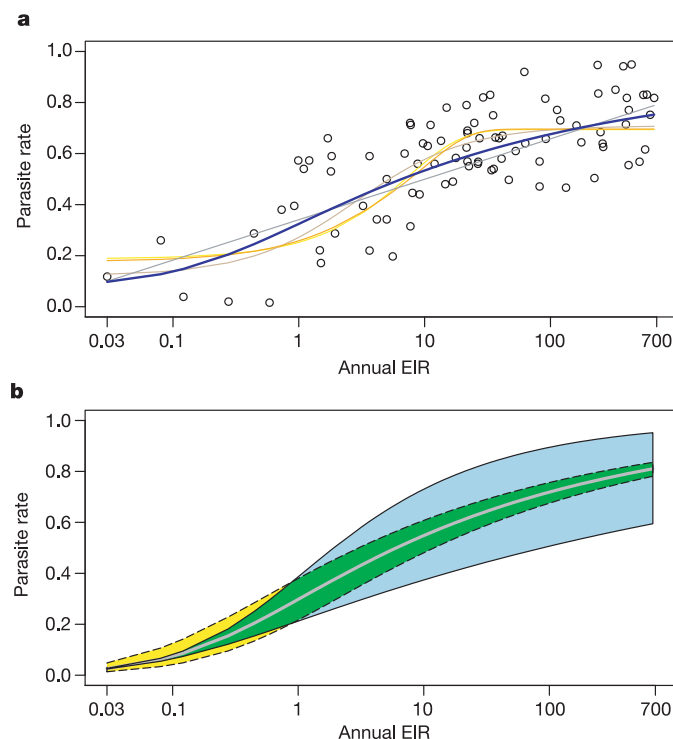


Figure 1 | The data and fitted functions and a sensitivity analysis for the best overall model. **a**, The best fit for SIS (tan), $\text{SI}^{\circ}\text{S}$ (yellow), $\text{SI}^{\circ}\text{RS}$ (orange), the log-linear model (grey), and the best-overall model $\int \text{SI}^{\circ}\text{S}$ (blue). The best-fit $\int \text{SI}^{\circ}\text{RS}$ model was not visually different from the SIS model with heterogeneity. **b**, In $\int \text{SI}^{\circ}\text{S}$, PR is more sensitive to the variance in infection rates (that is, doubling or halving k) at high EIR (solid lines, blue and green), and more sensitive to variability in recovery (doubling or halving r) for low EIR (dashed lines, yellow and green). Models are defined in Methods.

most of the effect was due to 16 studies for which the maximum age was larger than 12. One likely explanation is that these studies included many children who had become sufficiently immune to control the peripheral parasitaemia. Individuals who control peripheral parasitaemia may clear infections faster, or they may be more likely to return a false negative microscopy report, an issue that may be resolved with a more sensitive test such as PCR (polymerase chain reaction).

One objection to this analysis is that the heterogeneous composition of the population inevitably biases the study. The inferential perils of cross-level analysis can be avoided if the heterogeneous composition of the population is known in advance and incorporated into the analysis, but this is a practical impossibility. Our approach deals with this problem by assuming that the heterogeneity can be treated statistically. We specify a one-parameter family of distributions and allow a shape parameter to be fitted. This analysis corrects for the bias introduced by heterogeneity, assuming that the distribution is properly chosen, and that heterogeneous infection rates are the main source of bias.

We have found a simple approximating model for the relationship between EIR and PR. Our analysis suggests that heterogeneous biting or susceptibility to infection plays an important role in determining PR, that immunity to infection in early childhood does not, and that persistence times for malaria are at least six months, and possibly much longer.

Our findings have broad implications for malaria dynamics. Clearly, PR declines in adults^{28,29} owing to some kind of immunity that reduces infection, increases clearance, or that reduces apparent PR by lowering the parasite densities in peripheral parasitaemia and decreasing sensitivity by microscopy. The distinction between immunity to clinical disease, immunity to infection, transmission blocking immunity, and the ability of individuals to manage peripheral parasitaemia has been confusing in mathematical models; the epidemiological state often called 'recovered and immune' acts as a reservoir of *P. falciparum* in some models but not in others. A critical evaluation of the mathematical models of malaria is warranted.

The results here are consistent with earlier views of malaria epidemiology; very substantial reductions in EIR will be necessary to achieve modest reductions in PR throughout Africa. In quantitative terms, reducing EIR from 200 to 100 and then to 50 would reduce PR by 4% and then by an additional 5%. A corollary is that the effort required to reduce the disease burden in Africa would be enormous. Such findings have substantial implications for (1) the prospects of control with imperfect vaccines, (2) the development and persistence of anti-malarial drug resistance, and (3) public health interventions that aim to reduce disease without reducing PR. In particular, heterogeneous infection implies that malaria will be substantially more difficult to control if control measures are applied uniformly. Conversely, targeting malaria control at those who are bitten most, where practical, may provide a disproportionate impact and wider community benefits by reducing the frequency of asymptomatic infections, the sporozoite rate in the mosquito population and overall transmission¹⁻³.

METHODS

Models. Ross's population dynamic model describes the temporal relationship between EIR and PR, as well as the relationship at equilibrium¹². Let x denote PR in a population as a function of time or age. Initially, we assume people become susceptible to infections after clearing an infection. Assuming EIR remains constant, PR changes according to the following equation:

$$\dot{x} = b\mathcal{E}(1 - x) - rx \quad (2)$$

Ross's relationship between EIR and PR is found by setting $\dot{x} = 0$ and solving for x (Table 1). The model is called SIS, because those who are infected (I) become susceptible (S) again after clearing an infection.

For super-infection, PR is the equilibrium of the equation:

$$\dot{x} = b\mathcal{E}(1 - x) - f(b\mathcal{E}, r)x \quad (3)$$

where $f(\Lambda, r) = \Lambda/(e^{\Lambda/r} - 1)$. We call this an SI^oS model, where I^o implies the per capita time to clearance (that is, from I to S) is given by f .

In heterogeneous populations, let s index the population with expected infection rate $b\mathcal{E}$, and let $x(s)$ denote the proportion of humans in that class that are infected. To describe the distribution of infection rates in the population, let $g(s)$ denote the fraction of the population in class s , and without loss of generality, let $g(s)$ denote a probability distribution function with mean 1. Thus, $g(s)$ affects the distribution of infection rates without changing the mean; $b\mathcal{E}$ describes average infection rates, but individual expectations can vary substantially. The dynamics are described by the equation:

$$\dot{x}(s) = b\mathcal{E}(1 - x(s)) - f(b\mathcal{E}, r)x(s) \quad (4)$$

The population prevalence is found by solving for the equilibrium in equation (4), denoted $\bar{x}(s)$, and integrating:

$$\int_0^\infty \bar{x}(s)g(s)ds \quad (5)$$

Here, we let $g(s, k)$ denote a Γ distribution, with mean 1 and variance $1/k$. Thus, the average rate of infection in the population is $b\mathcal{E}$ and the variance of the infection rate is $b^2\mathcal{E}^2/k$; $1/k$ is the coefficient of variation of the population infection rate. For this distribution, equation (5) has the closed form solution given by equation (1). This model is called $\int$ SI^oS.

Ross's model, the heterogeneous infection model, and the superinfection model are closely related. As expected, the functional relationship with super-infection is the limit of a heterogeneous infection model as the variance in expected infection rates approaches 0. Curiously, Ross's original function is a special case of a heterogeneous infection model (equation (1)) with $k = 1$. A longer closed form expression can be derived for the model $\int$ SIS, the heterogeneous model with Ross's assumption about clearance (not shown). The best fit model $\int$ SI^oS is virtually identical to the Ross analogue of the best fit model $\int$ SIS but with a very different interpretation (results not shown). Thus, the super-infection clearance assumption does little, *per se*, to improve the model fit. On the other hand, it may provide a more accurate estimate of the time to clear an infection¹⁹.

For immunity to infection, let y denote the proportion of a population that has cleared *P. falciparum* infections and is immune to re-infection. Let ϕ denote the average duration of immunity to re-infection. The dynamics are described by the equations:

$$\begin{aligned} \dot{x} &= b\mathcal{E}(1 - x - y) - f(b\mathcal{E}, r)x \\ \dot{y} &= f(b\mathcal{E}, r)x - y/\phi \end{aligned} \quad (6)$$

Note that the fitted parameter is actually $\phi' = b\phi$ (see Table 1). This model is called SI^oRS, where R means recovered and immune.

For a heterogeneous population model with immunity to infection, let $y(s)$ denote the proportion of recovered and immune hosts. The dynamics are described by the equations:

$$\begin{aligned} \dot{x}(s) &= b\mathcal{E}(1 - x(s) - y(s)) - f(b\mathcal{E}, r)x(s) \\ \dot{y}(s) &= f(b\mathcal{E}, r)x(s) - y(s)/\phi \end{aligned} \quad (7)$$

We could not find a closed-form expression, so we fitted the function shown in Table 1; numerical integration was performed by R . This model is called $\int$ SI^oRS. **Age, microscopy errors and likelihood.** Let α denote the sensitivity of microscopy and $1 - \beta$ the specificity. The estimated PR, Y_i , is related to the true PR by the formula $Y = \alpha X + \beta(1 - X)$; it is biased upwards at low prevalence by false positives (β) and downwards at high prevalence by false negatives (α).

Similarly, the differences in the age distribution of children sampled is a potential source of bias. As we have no information about the age distribution of children actually sampled, we use the bounds for bias correction.

Let L_i and U_i be the lower and upper ages of the children from the i th study, and let $\bar{L}$ and $\bar{U}$ denote the mean lower and bounds from all the studies. The predicted PR in the i th sample, given the estimated EIR, is $X_i = f(\mathcal{E}_i) + a(L_i - \bar{L}) + b(U_i - \bar{U})$. The predicted estimate of PR in the i th sample, given the estimated EIR is $Y_i = \alpha X_i + \beta(1 - X_i)$. The log likelihood of the data, given the model, was:

$$\sum_{i=1}^{91} P_i \log(Y_i) + N_i \log(1 - Y_i) \quad (8)$$

where P_i was the number of humans that tested positive and N_i the number that tested negative in the i th study. For each underlying function f , we verified that the full model was in fact the best model by backwards fitting (that is, we redid the analysis setting α , β , a and b to zero, one by one, then in combinations).

We have reported the fitted sensitivity and specificity, but we also repeated the analysis assuming a constant sensitivity and specificity for all models. The results and model ranking are similar to those reported here if the assumed values of sensitivity and specificity are high, close to their fitted values. In contrast, when sensitivity and specificity are much lower, the log-linear model is strongly favoured. As the lowest PR that could be measured is β , and as the PR in several studies was close to zero, specificity must be fairly high, at least when PR is low. Some of this information is captured by fitting α and β .

Uncertainty in EIR. The studies used different methods to estimate EIR, so we were uncertain how to estimate the study precision. Moreover, only 38 studies reported sample sizes for their estimates of the human biting rate and the sporozoite rate. To evaluate whether uncertainty in EIR would affect our analysis, we repeated the analysis 1,000 times using equation (1) and the log-linear model. In each instance, we multiplied the estimated EIR by 2^ξ , where ξ was a uniform random variable drawn from the interval $(-1, 1)$. The average best-fit parameter estimates changed, but equation (1) was selected as the best model in 98.8% of these trials. We repeated the analysis again multiplying EIR by 4^ξ with similar results.

Received 26 May; accepted 11 July 2005.

- Dietz, K. Models for vector-borne parasitic diseases. *Lecture Notes Biomath.* **39**, 264–277 (1980).
- Dye, C. & Hasibeder, G. Population dynamics of mosquito-borne disease: effects of flies which bite some people more frequently than others. *Trans. Soc. Trop. Med. Hyg.* **80**, 69–77 (1986).
- Woolhouse, M. E. *et al.* Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342 (1997).
- Hay, S. I., Guerra, C. A., Tatem, A. J., Atkinson, P. M. & Snow, R. W. Urbanization, malaria transmission, and disease burden in Africa. *Nature Rev. Microbiol.* **3**, 81–90 (2005).
- Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M. & Snow, R. W. The global distribution and population at risk of malaria: past, present and future. *Lancet Infect. Dis.* **4**, 327–336 (2004).
- Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* **434**, 214–217 (2005).
- Murray, C. J. L. & Lopez, A. D. Alternative projections of mortality and disability by cause 1990–2020: Global burden of disease study. *Lancet* **24**, 1498–1504 (1997).
- Jamison, D. T., Creese, A. & Prentice, T. *World Health Report, 1999: Making a Difference* (World Health Organization, Geneva, 1999).
- Snow, R. W. & Gilles, H. M. in *Essential Malariology* 4th edn (eds Warrell, D. A. & Gilles, H. M.) Ch. 5, 85–106 (Arnold, London, 2002).
- Snow, R. W. & Marsh, K. The consequences of reducing transmission of *Plasmodium falciparum* in Africa. *Adv. Parasitol.* **52**, 235–264 (2002).
- Macdonald, G. *The Epidemiology and Control of Malaria* (Oxford Univ. Press, London, 1957).
- Ross, R. *The Prevention of Malaria* (John Murray, London, 1911).
- Nájera, J. A. A critical review of the field application of a mathematical model of malaria eradication. *Bull. World Health Organ.* **50**, 449–457 (1974).
- Smith, D. L., Dushoff, J. & McKenzie, F. E. The risk of a mosquito-borne infection in a heterogeneous environment. *PLoS Biol.* **2**, e368 (2004).
- Takken, W. & Knols, B. G. J. Odor-mediated behaviour of Afrotropical malaria mosquitoes. *Annu. Rev. Entomol.* **44**, 131–157 (1999).
- Hethcote, H. W. & Yorke, J. A. In *Lecture Notes in Biomathematics* Vol. 56, 1–105 (Springer, Berlin, 1984).
- Walton, G. A. On the control of Malaria in Freetown, Sierra Leone. I. *Plasmodium falciparum* and *Anopheles gambiae* in relation to malaria occurring in infants. *Ann. Trop. Med. Parasitol.* **41**, 380–407 (1947).
- Dietz, K., Molineaux, L. & Thomas, A. A malaria model tested in the African savannah. *Bull. World Health Organ.* **50**, 347–357 (1974).
- Aron, J. L. & May, R. M. in *Population Dynamics and Infectious Disease* (ed. Anderson, R. M.) Ch. 5, 139–179 (Chapman and Hall, London, 1982).
- Bailey, N. T. J. *The Biomathematics of Malaria* (Oxford Univ. Press, Oxford, 1982).
- McKenzie, F. E., Sirichaisinthop, J., Miller, R. S., Gasser, R. A. & Wongsrichanalai, C. Jr Dependence of malaria detection and species diagnosis by microscopy on parasite density. *Am. J. Trop. Med. Hyg.* **69**, 372–376 (2003).
- R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2004); (<http://www.R-project.org.3-900051-07-0>)
- Beier, J. C., Killeen, G. F. & Githure, J. I. Short report: entomologic inoculation rates and *Plasmodium falciparum* malaria prevalence in Africa. *Am. J. Trop. Med. Hyg.* **61**, 109–113 (1999).
- Burnham, K. P. & Anderson, D. A. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* 2nd edn (Springer, New York, 2002).
- Eyles, D. E. & Young, M. D. The duration of untreated or inadequately treated *Plasmodium falciparum* infections in the human host. *J. Nat. Malaria Soc.* **10**, 327–336 (1951).
- Gu, W. *et al.* Low recovery rates stabilize malaria endemicity in areas of low transmission in coastal Kenya. *Acta Trop.* **86**, 71–81 (2003).
- Sama, W., Killeen, G. & Smith, T. Estimating the duration of *Plasmodium falciparum* infection from trials of indoor residual spraying. *Am. J. Trop. Med. Hyg.* **70**, 625–634 (2004).
- Bagster Wilson, D. Rural hyperendemic malaria in Tanganyika Territory. *Trans. R. Soc. Trop. Med. Hyg.* **29**, 583–618 (1936).
- Molineaux, L. & Gramiccia, G. *The Garki Project* (World Health Organization, Geneva, 1980).

Acknowledgements We thank D. Bradley, B. Grenfell, F. E. McKenzie, W. Prudhomme, S. Randolph, M. Recker, D. Rogers and L. Waller for comments and suggestions. We also thank the NCEAS working group for discussion. S.I.H. is funded by a Research Career Development Fellowship from the Wellcome Trust. R.W.S. is a Wellcome Trust Senior Research Fellow and acknowledges the support of the Kenyan Medical Research Institute (KEMRI). This work was partially conducted as part of the Environment and Disease Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF, the University of California Santa Barbara, and the State of California. The views presented in this Letter represent the personal views of the authors and do not construe or imply any official position or policy of the Fogarty International Center, National Institutes of Health, Department of Health and Human Services, or the US government.

Author Contributions S.I.H. collated the data. D.L.S. designed and conducted the analysis. D.L.S., J.D., S.I.H. and R.W.S. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L.S. (smidave@helix.nih.gov).

LETTERS

Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus

John-Dylan Haynes^{1,2}, Ralf Deichmann¹ & Geraint Rees^{1,2}

When dissimilar images are presented to the two eyes, they compete for perceptual dominance so that each image is visible in turn for a few seconds while the other is suppressed. Such binocular rivalry is associated with relative suppression of local, eye-based representations^{1–4} that can also be modulated by high-level influences such as perceptual grouping^{3,5,6}. However, it is currently unclear how early in visual processing the suppression of eye-based signals can occur. Here we use high-resolution functional magnetic resonance imaging (fMRI) in conjunction with a new binocular rivalry stimulus to show that signals recorded from the human lateral geniculate nucleus (LGN) exhibit eye-specific suppression during rivalry. Regions of the LGN that show strong eye-preference independently show strongly reduced activity during binocular rivalry when the stimulus presented in their preferred eye is perceptually suppressed. The human LGN is thus the earliest stage of visual processing that reflects eye-specific dominance and suppression.

Currently, little is known about the site and mechanisms of eye-specific suppression during binocular rivalry. The earliest stage after the retina at which differential eye-specific modulation could occur is the LGN, which contains monocular neurons segregated into eye-specific layers⁷. Although excited only by monocular stimulation, neurons in primate LGN can nevertheless show robust inhibitory binocular interactions⁸. However, fluctuations in activity associated with changing perception during binocular rivalry have proven inconsistent in cat^{9,10} and are apparently absent in monkey¹¹. There have been no reported investigations of binocular rivalry in the human LGN. Common to all previous studies is the absence of behavioural measures of perceptual dominance, either because the animals were anaesthetized or because they were not required to report their perceptual experience. The lack of behavioural measurements for classifying neuronal responses will have significantly lowered the power of such studies to detect any neural signature of rivalry, as perceptual dominance must instead be inferred indirectly.

We therefore set out to investigate whether the human LGN shows eye-specific changes in signal in association with behaviourally measured changes in perception during binocular rivalry. First, we functionally identified the LGN independently in each of four participants by contrasting contralateral and ipsilateral hemifield binocular stimulation (Fig. 1a, see Supplementary Information). Next, we investigated responses of voxels within the LGN to purely monocular stimulation with a bilateral stimulus (Fig. 1b; see Methods). The human LGN comprises six histologically distinct monocular layers, each approximately 1-mm thick. Although the overall size of the human LGN can vary by a factor of two⁷, individual layers are still below the spatial resolution of conventional human fMRI.

To maximize sensitivity, we used an improved fMRI scanning sequence with twice the spatial resolution of conventional techniques

(1.5 mm isotropic; see Methods). Although this much higher spatial resolution might still be too low to permit direct visualization of each monocular layer, we hypothesized that the ocular preferences of neuronal populations within the LGN would be revealed through each voxel providing a biased sampling of monocular cells. Such biased sampling arises because a randomly placed voxel will sample an anisotropic distribution of ocular preferences, owing to the convoluted nature of the laminar structure of the LGN. Similar biases have been demonstrated for orientation preference in human V1, even at the lower spatial resolution of conventional fMRI, and have been used to directly measure orientation-selective processing^{12,13}.

Confirming our hypothesis of ocular biases in human LGN, voxels within the LGN responded vigorously to stimulation of either eye, but many showed a clear and significant preference or bias towards either right or left eye stimulation (Fig. 1c). For every voxel in the LGN in each participant, we independently computed a measure of this ocular bias. The distribution of these biases across all LGN voxels followed an approximately gaussian distribution (Supplementary Fig. S1). Between 42 and 71% of all LGN voxels in each participant showed a significant ($P < 0.05$) bias towards either ipsilateral or contralateral eye stimulation.

Having established the presence of significant ocular biases within LGN voxels, we could then proceed to test responses within these functionally defined voxels during binocular rivalry. We measured activity in the LGN and V1 using high-resolution fMRI while our participants viewed a novel binocular rivalry stimulus (see Fig. 2 and Methods) consisting of two co-rotating orthogonal gratings (Fig. 2a). This stimulus was specially devised to ensure relatively long perceptual dominance periods with minimal piecemeal rivalry, while strongly driving neuronal activity in the LGN and early visual cortex (Fig. 2a). Participants reported which stimulus was dominant by pressing a button. We confirmed that dominance phases followed a gamma distribution¹⁴, with a relatively long mean perceptual dominance of 10.7 s (Fig. 2b).

We next investigated whether voxels with left- or right-eye preference showed different responses when the left-eye or right-eye stimulus became dominant. For each LGN voxel, we calculated the fMRI signal associated with perceptual dominance and suppression of each monocular image. We then used the information about the ocular preference of each voxel (obtained from the participant's independent eye-localizer session) to calculate the fMRI signal associated with perceptual dominance and suppression of the stimulus presented to the preferred eye of each voxel. Signals during dominance of either left- or right-eye images were calculated separately for all voxels with biases towards either the left or right eye. This revealed strong and highly significant increases in the LGN signal during rivalry whenever the monocular stimulus corresponding to the independently measured ocular bias of that voxel became

¹Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, UK. ²Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, UK.

dominant, with corresponding decreases when it was suppressed (Fig. 3a). For example, voxels with left-eye preference under monocular viewing conditions (Supplementary Fig. S1) showed relative enhancement of signals under rivalrous viewing when the left eye stimulus became dominant (Fig. 3a, red symbols in right panel). Conversely, voxels with a preference for right-eye stimulation showed relative enhancement when the right-eye stimulus became perceptually dominant. These monocular signals of eye-specific enhancement and suppression reflecting perceptual dominance during rivalry were replicated independently in every participant (Supplementary Fig. S2). These time courses also closely followed the

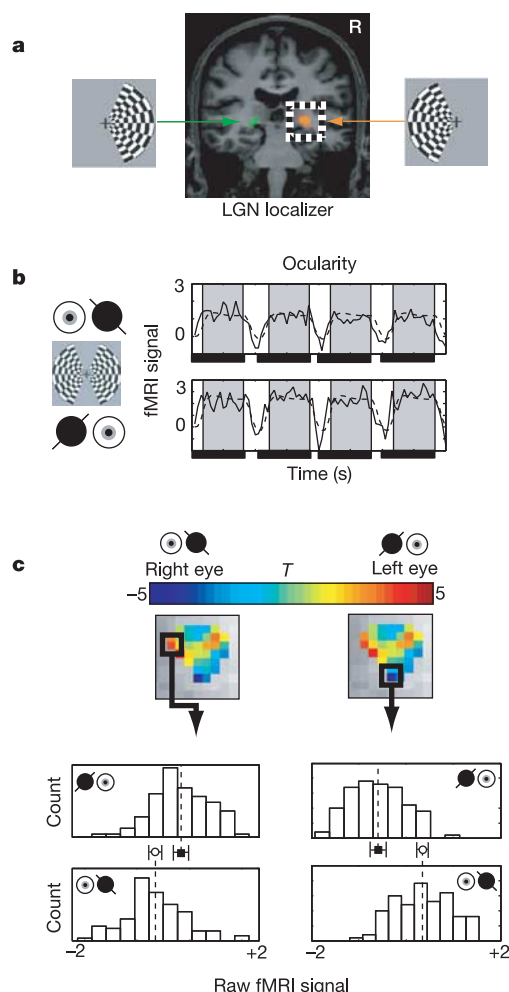


Figure 1 | Ocular biases in the LGN. **a**, The left and right LGN were localized for each participant as the regions in the left and right lateral posterior thalamus responding more strongly to contralateral than to ipsilateral hemifield stimulation³¹. **b**, Eye-specific stimulation. In alternate runs either the right eye alone (top) or the left eye alone (bottom) was stimulated with a double-wedge stimulus that was presented in blocks of 30 s (grey shaded regions). For each voxel in the LGN (see **a**), we extracted the fMRI signal amplitude during stimulation blocks for stimulation of each eye separately (shown here for one voxel in mean corrected scanner units). **c**, For each voxel, we then investigated whether the fMRI responses were stronger to stimulation of the left eye (positive *T*-values) or right eye (negative *T*-values). The coloured graphs show the colour-coded distribution of these ocular preferences for a coronal section through one participant's right LGN (dashed box in **a**), and the black and white histograms show the raw fMRI signal distribution for a left-eye-biased and a right-eye-biased voxel during periods when either the left or the right eye was stimulated (grey shaded periods in **b**). Although the raw signal distributions largely overlap, there was a significant bias for these two voxels as indicated by the means (±s.e.m.) for each distribution (shown with dotted lines, open circles and filled squares; see Supplementary Fig. S1 for full results).

independent predictions of a parameter-free 'forward model', obtained from the perceptual time courses by convolution with a haemodynamic response function (Fig. 3a, solid red and blue lines; see Methods).

To formally characterize the eye-dependence of relative suppression and enhancement, we calculated the difference in activity between periods of perceptual dominance of the left-eye stimulus (red) and the right-eye stimulus (blue) across all voxels dominated by either the left eye (Fig. 4; red bars) or the right eye (Fig. 4; blue bars). Activity in left-eye-dominated voxels was significantly stronger when the left-eye stimulus (red) was dominant than when the right-eye stimulus (blue) was dominant. This is consistent with the notion that binocular rivalry involves differential modulation of eye-specific signals during perceptual dominance and suppression. To estimate the size of this modulation, we compared the signal change associated with eye-specific modulation during rivalry with the signal change evoked by extended monocular viewing of each monocular stimulus in one representative participant (see Supplementary Fig. S3). Although great caution is required in comparing these two very different measures in structures with non-specific monocular inhibitory interactions such as the LGN, eye-specific modulation during rivalry was incomplete compared with purely monocular stimulation (64% modulation; see Methods).

Having shown that LGN voxels with a specific ocular preference showed signals related to perceptual dominance of that eye during rivalry, we next determined whether there was a graded relationship between these two independent measures. For each LGN voxel, we plotted the difference in activation between perceptual dominance of the preferred-eye and the non-preferred-eye stimulus as a function of its ocular preference (Supplementary Fig. S4). There was a monotonic relationship between these two measures, with a modest but statistically significant positive correlation ($P < 0.05$) between the two for every participant. Thus, voxels with relatively strong ocular biases tended to show larger differences in activity during different dominance phases in rivalry. This might arise because signals in voxels with strong ocular biases are dominated by the input mainly from one eye and thus more clearly reflect the difference between rivalry and suppression.

Finally, we used the same methodology to quantify ocular biases and differential activity during perceptual dominance in rivalry in V1 (see Methods). As in the LGN, there was a very similar relationship between the ocular preference of individual voxels in V1 and the presence of differential signals during perceptual dominance during rivalry (Figs 3 and 4). Eye-specific modulation during rivalry was also incomplete compared with purely monocular stimulation (28% modulation; see Supplementary Fig. S3). Small quantitative

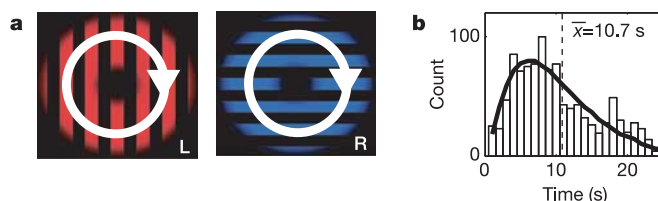


Figure 2 | Binocular rivalry stimulus. **a**, A red grating was presented to the left eye and an orthogonal blue grating was presented to the right eye while both stimuli simultaneously rotated clockwise. They were viewed through red/blue anaglyph glasses, so that the blue grating was visible to the right eye and the red grating was visible to the left eye. During each scanning run, the two stimuli were presented continuously for 5 min, leading to multiple perceptual alternations between the two monocular images while the other image was suppressed. Participants used response buttons to continuously indicate during each run whether they currently perceived either the red or blue grating. **b**, Histogram of dominance phase durations averaged across participants, which reveals that dominance phases were relatively long compared with typical rivalry stimuli, and could be well fitted by a gamma-distribution function (black line).

differences (for example, smaller overall differential activity; compare ordinate axes of Fig. 3a and b) might reflect differences in baseline fMRI signals, in the width of the haemodynamic point-spread function, or in the response gain between LGN and V1 (refs 15, 16).

These findings represent positive evidence for a neural signature of rivalry in the LGN of a conscious individual of any species. Rivalry-related signals are thus not confined to the early visual cortex^{17–20}, but can also be observed in subcortical structures. Using single-unit electrophysiological recording in cats and monkeys, it has proven difficult to establish whether the LGN reflects perceptual dominance in rivalry^{9–11}. These studies used different stimuli (for example, presented at greater eccentricities¹¹) in different species and did not measure perceptual dominance behaviourally, making it unclear at which point during recording any perceptual transitions might have occurred. Such studies thus require a substantially different analytic approach compared to the present study. For example, one study in monkeys¹¹ used power spectral analysis of neural signals, which should in principle allow accumulation of transition-related effects even though their random phase position is unknown. However, this approach is less sensitive than the one taken here, where the times of perceptual transitions are known, allowing for event-related averaging. Our successful demonstration of rivalrous fluctuations in the human LGN might in principle reflect any of these major differences. Alternatively, it might reflect differential sensitivity of the fMRI technique to particular aspects of the neuronal signal (for example, local field potentials rather than spiking activity²¹).

Our findings indicate that dominance and suppression in rivalry are accompanied by a modulation of signals in eye-specific representations in the LGN and V1. This is consistent with previous observations of neural signatures of rivalry in the monocular representation of the blind spot in V1 (ref. 17). However, by demonstrating that individual voxels acquired with high-resolution fMRI show reliable ocular biases, we were able to show that such

monocular rivalrous signatures are not specific to the unusual cortical locale of the blind spot but extend throughout V1 and into the LGN (Figs 3 and 4). The suggestion²² that previous findings of rivalry in the blind-spot representation¹⁷ in V1 might reflect competition between different stimulus preferences in surrounding binocular cells cannot account for our findings in LGN, which has no binocular cells.

Although binocular rivalry can prevent conscious perception of a suppressed image, visual sensitivity during suppression is only moderately decreased^{23,24}. Moreover, suppressed images can still cause simple after-effects^{25,26}. This suggests that the eye-specific modulation we observed in human LGN must necessarily be incomplete, consistent with recent reports that higher visual areas show selective responses to suppressed stimuli²⁷. Similarly, we found that the depth of rivalrous modulation was smaller than the signal evoked by purely monocular stimulation with our stimuli. It is important to note that any differences between monocular stimulation and rivalry are only approximations of modulation depth. Such differences potentially reflect not only rivalry-related signal modulations, but also the absence of non-specific inhibitory interactions under monocular stimulation⁸. Furthermore, differences in the degree of suppression between different structures (here, LGN and V1) may be influenced by differences in response gain^{15,16}. These caveats notwithstanding, our findings are consistent with the notion that suppression during rivalry does not result in a complete inhibition of stimulus-driven signals, and that sensitivity to input from the suppressed eye is moderately (but not fully) reduced^{23,24}. In contrast, selective adaptation by suppressed images can be of equal magnitude as for dominant images²⁶, suggesting that visual responses may reach early visual areas largely unattenuated; but adaptation is moderately reduced when visual saturation of the adapting stimuli is avoided²⁵.

Responses to flashed stimuli in the LGN can show binocular interactions earlier than the first V1 responses⁸, suggesting that such interactions do not require corticofugal feedback. Here, stimuli were presented continuously for several minutes, resulting in a superposition of feed-forward and feedback processing owing to the dense reciprocal connections between V1 and the LGN²⁸. Thus, our data do not distinguish the precise source of the modulatory influence that we observed in LGN. Because our stimuli have rivaling orientations, eye-specific modulations in LGN could potentially reflect a feedback signal from V1, where orientation-selective processing is well established^{12,13}. This is consistent with BOLD signals correlating slightly better with local field potentials than with spikes²¹. If feedback plays such a role then it must necessarily be eye-specific, thus placing specific constraints on the possible underlying neural circuitry²⁸. Alternatively, LGN modulation may instead arise directly from binocular interactions between LGN layers^{8,29} through local inhibitory interneurons³⁰. Such local computations should also be visible in the BOLD response²¹. Note that our findings do not imply that rivalry is exclusively determined by eye-specific

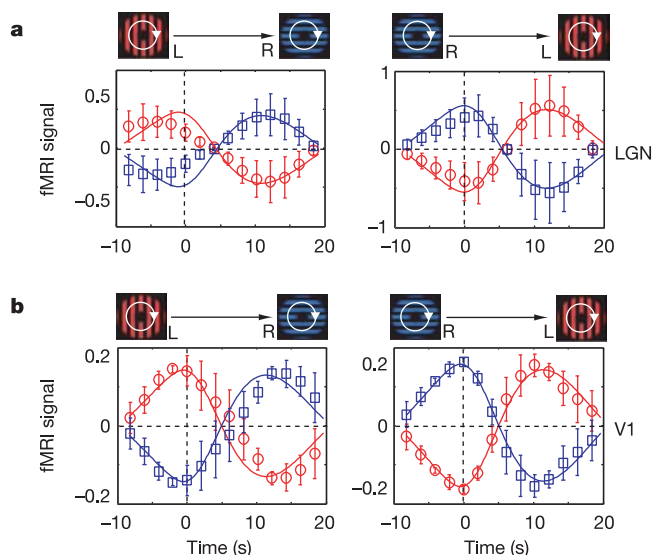


Figure 3 | Rivalry-related responses in LGN and V1. **a**, Event-related BOLD-fMRI signal changes in the LGN averaged across participants, time-locked to perceptual transitions from red to blue (left) and from blue to red (right). Error bars represent standard error across participants. Event-related averages were computed separately for voxels with left-eye preference (red circles) and right-eye preference (blue squares) as determined in the independent localizer sessions. The solid red and blue lines show the prediction of a parameter-free forward model of the haemodynamic response to perceptual transitions reported by the participants (see Methods). A good fit between behaviourally predicted and observed haemodynamic responses is apparent. **b**, As in **a**, but for voxels in the primary visual cortex (V1).

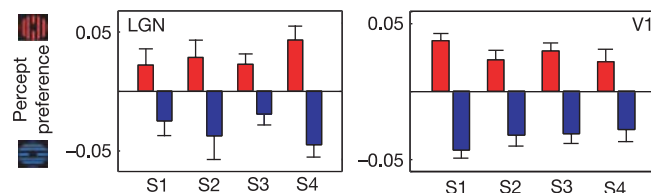


Figure 4 | Percept preference in LGN and V1. A 'percept preference' measure in the LGN (left) and V1 (right) was obtained separately for left-eye-dominated voxels (red bars) and right-eye-dominated voxels (blue bars) by subtracting the average BOLD signal during blue-dominance from the average BOLD signal during red-dominance (see Supplementary Information). Positive values indicate that signal was higher during dominance of the red percept, and negative values that signal was higher during dominance of the blue percept. Data are for subjects S1–S4; error bars represent standard error across voxels).

representations. Coherence between global stimulus-representations can clearly have a role in resolving rivalry^{3,5}. Such global effects might operate through local eye-based modulation in the LGN (or V1), modulated by top-down biases that favour coherent global percepts³. One possible control signal for such modulation could be feedback projections to monocular regions of V1 or even individual eye-specific layers of the LGN²⁸. Thus, rivalry may involve dynamic and competitive multilevel interactions at multiple processing stages¹.

METHODS

Participants and experimental design. Four healthy volunteers with normal vision (29–34 years old) gave written informed consent to participate in the experiment, which was approved by the local ethics committee. One participant was an author, two were experienced psychophysical observers and one was a naive participant.

The experiment consisted of three independent stages. In the first stage ('LGN localizer') we localized the left and right LGN in each participant using standard definitions^{15,31} (see Supplementary Information). The second stage ('eye localizer') sought to identify any ocular preference for individual LGN voxels. Participants fixated centrally while viewing a large contrast-reversing checkerboard stimulus presented bilaterally as two wedges (Fig. 1b), in alternating runs either to the left or right eye alone, while the other eye was covered and thus unstimulated. In the third stage we studied responses in the LGN and V1 to rotating binocular rivalry stimuli presented continuously for 296 s (see Fig. 2a and Supplementary Information). The stimuli yielded very effective rivalry with long phase durations (mean 10.7 s; see Fig. 2b) and very brief transition periods. Piecemeal or patchy rivalry was rare and very brief, and corresponding phases were discarded from the analysis.

fMRI acquisition. A 3T Allegra head scanner (Siemens Medical Systems) with a standard transmit–receive head coil was used to acquire functional data with a single-shot gradient echo isotropic high-resolution echo planar imaging sequence (matrix size 128 × 128; field of view 192 mm; in-plane resolution 1.5 mm; 20 slices with interleaved acquisition; slice thickness 1.5 mm; echo time 30 ms; acquisition time per slice 102 ms; repetition time 2,040 ms; echo spacing 560 μs; receiver bandwidth 250 kHz; 30% ramp sampling; twofold read oversampling to allow for *k*-space re-gridding; read gradient amplitude 34.47 mT m⁻¹; read gradient slew rate 344.7 mT m⁻¹ ms⁻¹). In order to maximize statistical power, we used only 20 slices that were optimized to cover the entire LGN, slightly angled to also achieve coverage of the calcarine sulcus.

In the main rivalry experiment, between 7 and 8 runs of 145 functional MRI volumes were acquired per participant. For the LGN localizer we acquired one run with 203 volumes, and for the eye localizer we acquired two runs of 163 volumes. For each participant a T1-weighted structural image was also acquired as well as 2–3 retinotopic mapping runs of 165 volumes, during which participants viewed standard flickering checkerboard stimuli that stimulated either the horizontal or vertical meridians.

fMRI analysis. Each participant's left and right LGN and V1 were identified using standard methods. An index of eye-dominance was then computed for each voxel in LGN and V1 (see Supplementary Information). To study the effects of rivalry on responses, we then investigated whether any significant response differences could be found between left-eye- and right-eye-dominated voxels during left-eye (red) and right-eye (blue) dominance phases. First we computed event-related responses (Fig. 3a, b and Supplementary Fig. S2) by selectively averaging the BOLD fMRI signal separately for left-eye-dominated and right-eye-dominated voxels, time-locked to changes in perceptual dominance either from red to blue or from blue to red. Error bars reflect standard errors across participants (Fig. 3a, b) or standard errors across all repetitions of one event-type for one participant (Supplementary Fig. S2). A parameter-free forward model of event-related responses was computed by selectively averaging a predicted fMRI time course obtained by convolving the perceptual time series with a canonical haemodynamic response function provided by the statistics package SPM2 (solid lines in Fig. 3). For a more formal comparison (Fig. 4) we also modelled responses separately for each voxel, using a general linear model with a regressor that modelled the difference in response amplitude between left-eye and right-eye perceptual dominance (see Supplementary Information).

Received 10 May; accepted 22 August 2005.

Published online 23 October 2005.

- Blake, R. & Logothetis, N. K. Visual competition. *Nature Rev. Neurosci.* **3**, 13–21 (2002).
- Blake, R., Westendorf, D. & Overton, R. What is suppressed during binocular rivalry? *Perception* **9**, 223–231 (1980).

- Lee, S. & Blake, R. A fresh look at interocular grouping during binocular rivalry. *Vision Res.* **44**, 983–991 (2004).
- Lehky, S. R. An astable multivibrator model of binocular rivalry. *Perception* **17**, 215–228 (1988).
- Kovacs, I., Papathomas, T. V., Yang, M. & Feher, A. When the brain changes its mind: interocular grouping during binocular rivalry. *Proc. Natl Acad. Sci. USA* **93**, 15508–15511 (1996).
- Logothetis, N. K., Leopold, D. A. & Sheinberg, D. L. What is rivaling during binocular rivalry? *Nature* **380**, 621–624 (1996).
- Andrews, T. J., Halpern, S. D. & Purves, D. Correlated size variations in human visual cortex, lateral geniculate nucleus, and optic tract. *J. Neurosci.* **17**, 2859–2868 (1997).
- Schroeder, C. E., Tenke, C. E., Arezzo, J. C. & Vaughan, H. G. Binocularity in the lateral geniculate nucleus of the alert monkey. *Brain Res.* **521**, 303–310 (1990).
- Varela, F. J. & Singer, W. Neuronal dynamics in the visual corticothalamic pathway revealed through binocular rivalry. *Exp. Brain Res.* **66**, 10–20 (1987).
- Sengpiel, F., Blakemore, C. & Harrad, R. Interocular suppression in the primary visual cortex: A possible neural basis of binocular rivalry. *Vision Res.* **35**, 179–195 (1995).
- Lehky, S. R. & Maunsell, J. H. R. No binocular rivalry in the LGN of alert macaque monkeys. *Vision Res.* **36**, 1225–1234 (1996).
- Haynes, J. D. & Rees, G. Predicting the orientation of invisible stimuli from activity in primary visual cortex. *Nature Neurosci.* **8**, 686–691 (2005).
- Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nature Neurosci.* **8**, 679–685 (2005).
- Levelt, W. J. Note on the distribution of dominance times in binocular rivalry. *Br. J. Psychol.* **58**, 143–145 (1967).
- Schneider, K. A., Richter, M. C. & Kastner, S. Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: a high-resolution functional magnetic resonance imaging study. *J. Neurosci.* **24**, 8975–8985 (2004).
- Sclar, G., Maunsell, J. H. & Lennie, P. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res.* **31**, 1148–1157 (1990).
- Tong, F. & Engel, S. A. Interocular rivalry revealed in the human cortical blind-spot representation. *Nature* **411**, 195–199 (2001).
- Polonsky, A., Blake, R., Braun, J. & Heeger, D. J. Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neurosci.* **3**, 1153–1159 (2000).
- Lee, S. H., Blake, R. & Heeger, D. J. Traveling waves of activity in primary visual cortex during binocular rivalry. *Nature Neurosci.* **8**, 22–23 (2005).
- Lee, S. H. & Blake, R. V1 activity is reduced during binocular rivalry. *J. Vis.* **2**, 618–626 (2002).
- Logothetis, N. K. & Wandell, B. A. Interpreting the BOLD signal. *Annu. Rev. Physiol.* **66**, 735–769 (2004).
- Andrews, T. J. Binocular rivalry and visual awareness. *Trends Cogn. Sci.* **5**, 407–409 (2001).
- Wales, R. & Fox, R. Increment detection thresholds during binocular rivalry suppression. *Percept. Psychophys.* **8**, 827–835 (1970).
- Watanabe, K., Paik, Y. & Blake, R. Preserved gain control for luminance contrast during binocular rivalry suppression. *Vision Res.* **44**, 3065–3071 (2004).
- Sobel, K. V., Blake, R. & Raissian, T. A. Binocular rivalry suppression does not impede buildup of the motion aftereffect. *J. Vis.* **4**, Abstract 243 (2004).
- Blake, R. & Fox, R. Adaptation to invisible gratings and the site of binocular rivalry suppression. *Nature* **249**, 488–490 (1974).
- Fang, F. & He, S. Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neurosci.* advance online publication, 4 September 2005 (doi:10.1038/nm1537).
- Ichida, J. M. & Casagrande, V. A. Organization of the feedback pathway from striate cortex (V1) to the lateral geniculate nucleus (LGN) in the owl monkey (*Aotus trivirgatus*). *J. Comp. Neurol.* **454**, 272–283 (2002).
- Singer, W. Control of thalamic transmission by corticofugal and ascending reticular pathways in the visual system. *Physiol. Rev.* **57**, 386–420 (1977).
- Montero, V. & Zempel, J. The proportion and size of GABA-immunoreactive neurons in magnocellular and parvocellular layers of the lateral geniculate nucleus of the monkey. *Exp. Brain Res.* **62**, 215–223 (1986).
- Kastner, S. et al. Functional imaging of the human lateral geniculate nucleus and pulvinar. *J. Neurophysiol.* **91**, 438–448 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Wellcome Trust. We thank J. Driver and P. Sterzer for comments, and E. Freeman for advice regarding the stimuli.

Author Contributions J.-D.H. and G.R. conceived the experiment, R.D. wrote the pulse sequence and J.-D.H. carried out the experiment and data analysis. G.R., J.-D.H. and R.D. co-wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.-D.H. (haynes@fil.ion.ucl.ac.uk).

LETTERS

Neural measures reveal individual differences in controlling access to working memory

Edward K. Vogel¹, Andrew W. McCollough¹ & Maro G. Machizawa¹

The capacity of visual short-term memory is highly limited, maintaining only three to four objects simultaneously^{1,2}. This extreme limitation necessitates efficient mechanisms to select only the most relevant objects from the immediate environment to be represented in memory and to restrict irrelevant items from consuming capacity^{3–5}. Here we report a neurophysiological measure of this memory selection mechanism in humans that gauges an individual's efficiency at excluding irrelevant items from being stored in memory. By examining the moment-by-moment contents of visual memory⁶, we observe that selection efficiency varies substantially across individuals and is strongly predicted by the particular memory capacity of each person. Specifically, high capacity individuals are much more efficient at representing only the relevant items than are low capacity individuals, who inefficiently encode and maintain information about the irrelevant items present in the display. These results provide evidence that under many circumstances low capacity individuals may actually store more information in memory than high capacity individuals. Indeed, this ancillary allocation of memory capacity to irrelevant objects may be a primary source of putative differences in overall storage capacity.

To examine the selection mechanism for allocating memory capacity, we recorded event-related potentials from healthy young adults while they performed a visual memory task⁷ in which it was necessary to remember selectively only a few relevant items from within an array. On each trial they were presented with a brief bilateral array of coloured rectangles of varying orientations and were asked to remember the orientations of only the red items in either the left or right hemifield, as indicated by an arrow (Fig. 1a). Memory for these red items was tested 1 s later with a test array that was either identical to the original memory array or differed by one orientation. Subjects reported whether the red items in the two arrays were identical or not by pressing one of two buttons. On a third of the trials, two red items were presented along with two blue items in each hemifield. On the remaining trials, arrays of either two red items or four red items alone were presented in each hemifield.

To observe directly whether the subjects could exclude the irrelevant blue items from being stored and maintained in visual memory, we measured a waveform of the event-related potential that reflects the encoding and maintenance of item representations in visual working memory⁶. This wave is a sustained negative voltage over the hemisphere that is contralateral to the memorized hemifield, and this activity persists throughout the memory retention interval. The amplitude of this contralateral delay activity (CDA) increases significantly as the number of representations being held in memory increases, reaching an asymptotic limit at each individual's specific memory capacity (ref. 6; A.W.M., M.G.M. and E.K.V., submitted). This limit is measured as a difference in amplitude between an array of four items and an array of two items. Low capacity individuals

show a smaller difference than high capacity individuals, indicating that an array of two items consumes a larger proportion of available memory capacity for low capacity subjects.

Because of the sensitivity of this measurement to the number of items that are currently held in memory, we used the CDA as a direct neurophysiological measure of whether or not the irrelevant distractor items unnecessarily consumed memory capacity. For example, on the trials in which two red items were presented simultaneously with two blue items, if an individual was perfectly efficient at remembering only the red items and excluding the blue items from memory, then the CDA amplitude should be equivalent to that observed when two red items were presented alone. By contrast, if an individual was perfectly inefficient at excluding the blue items, all four of the items in the array (two red and two blue) would be stored in memory, resulting in an amplitude equal to that when four red items alone were presented.

Memory capacity varies considerably across individuals, ranging

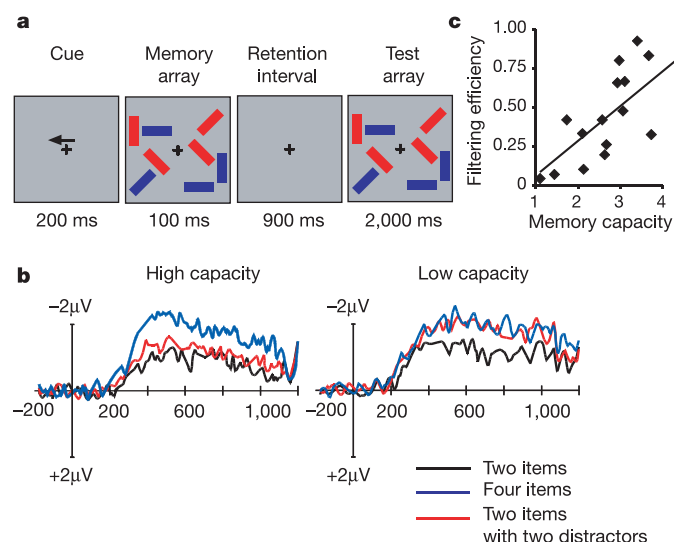


Figure 1 | Stimuli and results from experiment 1. **a**, Example of a 'distractors-present' trial for the left hemifield. **b**, Grand averaged ERP difference waves (contralateral activity minus ipsilateral activity) time-locked to the memory array averaged across the lateral occipital and posterior parietal electrode sites and divided across the high and low memory capacity groups. No significant differences in the pattern of effects were observed across the parietal and occipital electrode sites ($P > 0.30$). By convention, negative voltage is plotted upwards. **c**, Correlation between an individual's memory capacity and the efficiency of excluding distractors from being stored in visual working memory.

¹Department of Psychology, University of Oregon, Eugene, Oregon 97403-1227, USA.

from 1.5 objects to about 5 objects^{6,7}. To examine whether memory selection efficiency varies across memory capacity, we estimated each individual's memory capacity^{8,9} and divided the subjects into two groups: high and low capacity. These two groups differed markedly in their filtering efficiency abilities (Fig. 1b). For the high capacity group, the amplitude of the distractors-present condition was significantly smaller than that of four red items alone ($P < 0.001$) but was not significantly different from that of two red items alone ($P > 0.20$), indicating that these subjects were very efficient at excluding the distractors from consuming memory capacity. By contrast, the low capacity group had an amplitude in the distractors-present condition that was significantly larger than that in the two items alone condition ($P < 0.001$), but not significantly different from that in the four items alone condition ($P > 0.25$). These results indicate that low capacity subjects were highly inefficient at keeping the irrelevant items from being stored in memory.

We measured this relationship more formally by quantifying each subject's filtering efficiency (Methods). The scores are plotted as a function of each individual's memory capacity in Fig. 1c. These two measures were very strongly correlated ($r = 0.69$; $P < 0.001$): low capacity subjects showed low filtering efficiency scores, and high capacity subjects produced much higher efficiency scores. These results contrast with studies that have examined the neural bases of individual differences and have often reported complex relationships between the difficulty of the task and the magnitude of the neural activity^{10–12}. However, the CDA is primarily modulated by the number of objects held in memory rather than the difficulty of the task⁶, which may explain the simple relationship observed here.

The first experiment indicated that low capacity subjects are highly inefficient at excluding information on the basis of the colour of an item. However, previous research has shown that colour-based selection tends to be very difficult and inefficient relative to other selection attributes¹³. Consequently, it is possible that the relationship between memory capacity and filtering efficiency is present only under challenging filtering conditions. In experiment 2, we examined whether this relationship would generalize to a task in which subjects must filter distractors on the basis of location, a selection attribute that is considerably easier than selection by colour¹⁴. Figure 2a shows

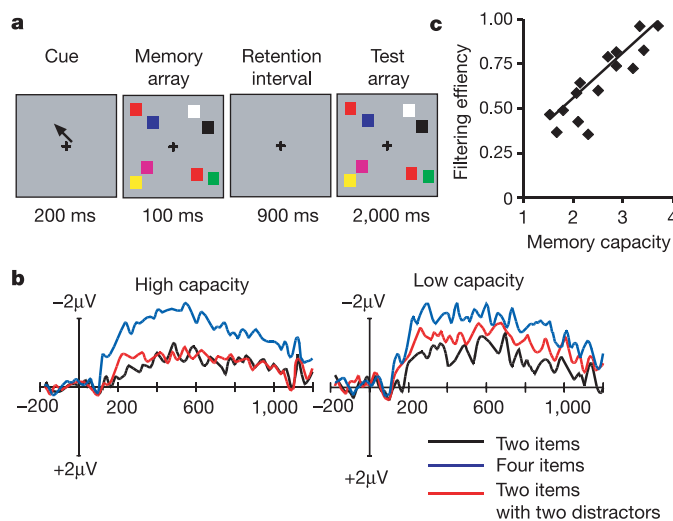


Figure 2 | Stimuli and results from experiment 2. **a**, Example of a 'distractors-present' trial for the upper portion of the left hemifield. **b**, Grand averaged ERP difference waves time-locked to the memory array averaged across the lateral occipital and posterior parietal electrode sites and divided across the high and low memory capacity groups. **c**, Correlation between an individual's memory capacity and the efficiency of excluding distractors from being stored in visual working memory.

an example of a distractors-present trial in which the subject is cued to remember the colours of only the items in the upper left quadrant and to exclude the items in the lower quadrant.

Figure 2b shows the CDA difference waves for the high capacity and low capacity groups for the three conditions. As in experiment 1, for the high capacity group the distractors-present condition had an amplitude that was equivalent to that in the two items alone condition and significantly smaller than in the four items alone condition ($P < 0.001$). By contrast, the low capacity group in the distractors-present condition had an amplitude that was significantly lower than in the four items alone condition ($P < 0.01$), but was significantly higher than in the two items alone condition ($P < 0.01$), indicating that this group was inefficiently storing information about some of the irrelevant distractors. Figure 2c shows the filtering efficiency scores plotted as a function of each subject's memory capacity. Whereas low capacity subjects were much more efficient than they were in the colour-based selection task, they were still considerably less efficient than the high capacity subjects ($r = 0.62$; $P < 0.001$), indicating that the relationship between memory capacity and filtering efficiency generalizes to both feature- and location-based selection.

It is plausible that the results of the first two experiments are due to a general inability of low capacity individuals to exert effective control over any aspect of working memory functioning, rather than to a more specific inability to exclude irrelevant items from being stored. An aspect of control over working memory is the ability to append new items into memory without overwriting existing items held in memory^{15,16}. In experiment 3, we examined whether low capacity subjects were also limited in their ability to append successfully, or whether their limitations are primarily restricted to situations that require the exclusion of irrelevant items from memory. Subjects were instructed to remember the orientations of only the red items in the cued hemifield. On half of the trials, subjects were presented with a single memory array that consisted of either two or four red items alone in the hemifield. On the other half of trials, subjects were presented a sequence of two memory arrays separated by 500 ms. The first memory array consisted of two red items and the second array consisted of either two red items which were to be appended (append red) or two green items which were to be excluded (exclude green). After a 1-s retention interval, all four items from both memory arrays were presented together in the test array and the subjects responded whether any of the red items had changed orientation or not.

Figure 3 shows the results of experiment 3 divided across high and low memory capacity subjects. For both groups, in the append red condition CDA amplitude was initially equivalent to an array size of

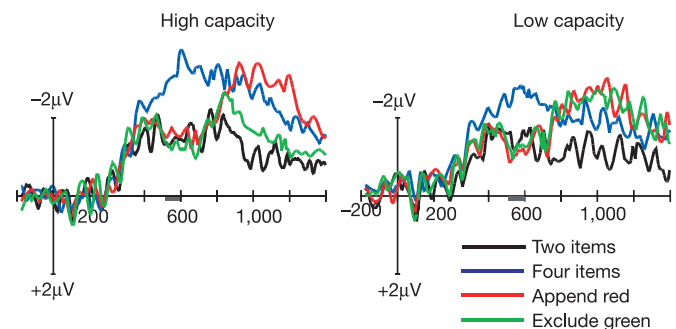


Figure 3 | Results from experiment 3. Shown are ERP difference waves at lateral occipital and posterior parietal electrode sites divided across high and low memory capacity groups. Grey rectangle indicates the duration for which the second array of items was present on the screen. As in experiments 1 and 2, low capacity individuals were less efficient at excluding the green items than were high capacity individuals ($r = 0.57$; $P < 0.01$).

two. Shortly after the onset of the second array, however, the amplitude rose to the equivalent of an array of four items. That is, both high and low capacity subjects showed a perfect additivity of the two arrays in terms of the amount of memory capacity consumed, indicating that low capacity subjects were not impaired at appending the items into working memory.

By contrast, in the exclude green condition large differences between the high and low capacity subjects were observed. For the high capacity subjects, CDA amplitude was initially equivalent to an array of two items. After the onset of the second array, CDA amplitude briefly rose to almost a four-item level but then quickly returned to near its original two-item level. For the low capacity subjects, however, CDA amplitude rose to the equivalent of a four-item level that was maintained throughout the retention interval, indicating that the green items had been unnecessarily appended into working memory. These results suggest that, although both low and high capacity subjects can append items into working memory, these two groups substantially diverge in their abilities to determine selectively which items will be appended into memory.

The control processes that regulate access to working memory are crucial for keeping irrelevant information from consuming capacity. Our results show that there is systematic variability across human individuals in the ability to control what is stored in working memory at any given moment. Neurophysiological studies in monkeys have indicated that the prefrontal cortex has a crucial role in determining what information is to be maintained in memory^{3,17,18}, and it is plausible that the individual differences reported in this study may stem from variability in a bias signal emanating from prefrontal cortex^{19–22}. A further implication of our study is that individual differences in memory capacity may not simply reflect variability in available storage space, but may also be strongly constrained by the efficiency with which the available space is allocated. By this view, an individual's specific memory capacity does not simply reflect 'how many' items can be stored, but also 'how efficient' the individual is at excluding irrelevant information from reaching this highly limited memory system²³.

METHODS

Subjects and experiments. Fifteen neurologically normal college students participated in each experiment (age range 19–28 yr) and gave informed consent according to procedures approved by the University of Oregon. Each of these observers performed between 200 and 240 trials per condition in each experiment. All stimulus arrays were presented within two $4^\circ \times 7.3^\circ$ rectangular regions that were centred 3° to the left and right of a central fixation cross on a grey background (8.2 cd m^{-2}). Stimulus positions were randomized on each trial, with the constraint that the distance between objects within a hemifield was at least 2° (centre to centre).

In experiment 1, each memory array consisted of two or four oriented rectangles ($0.65^\circ \times 0.65^\circ$) in each hemifield selected randomly from a set of four orientations (vertical, horizontal, left 45° and right 45°). In experiment 2, each memory array consisted of either two or four coloured squares in each hemifield. Each colour was randomly selected with limited replacement from a set of seven easily distinguished colours (red, blue, green, violet, yellow, black and white). The positions of the items were randomly distributed within the upper and lower quadrants of each hemifield. In the two items alone condition, both squares were presented in either the upper or the lower quadrant. In the four items alone condition, two items were presented in each quadrant.

In experiment 3, the first memory array consisted of either two or four red oriented rectangles. On half of the trials, a second memory array was presented 500-ms later consisting of two rectangles that were either red or green and were presented at new locations in the same general region as the first memory array. The 500-ms delay enables us to establish the CDA amplitude for the first memory array before the onset of the second array and provides sufficient time to extend beyond the duration of iconic memory for the first array.

Memory capacity and filtering efficiency. We computed visual memory capacity with a standard formula^{8,9} that essentially assumes that if an observer can hold in memory K items from an array of S items, then the item that changed should be one of the items being held in memory on K/S trials, leading to correct performance on K/S of the trials on which an item changed. To correct for guessing, this procedure also takes into account the false alarm rate. The formula

is $K = S(H - F)$, where K is the memory capacity, S is the size of the array, H is the observed hit rate, and F is the false alarm rate. Subjects were divided into high capacity and low capacity groups using a median split of their memory capacity estimates.

We quantified each individual's filtering efficiency with a formula in which we computed the mean amplitudes of the CDA across three conditions: two items alone, four items alone, and the distractors-present condition. In essence, this efficiency score measures whether the CDA amplitude on the distractors-present condition is more similar to that on the four items condition or the two items condition, with a range of scores from 1 (efficient: identical to two items) to 0 (inefficient: identical to four items). The formula is $\alpha = (F - D)/(F - T)$, where α is the filtering efficiency, F is the amplitude for four items, D is the amplitude for the distractors-present condition, and T is the amplitude for the two items alone condition. It is important to note that it is mathematically possible for this formula to yield a value outside the range 0 to 1 (for example, if $T > F$ or if $D < T$). Across all of the subjects in this study, however, there was no single case that met any of these conditions.

Electrophysiological recordings. Event-related potentials (ERPs) were recorded in each experiment using our standard recording and analysis procedures²⁴, including rejection of trials contaminated by blinks or large ($>1^\circ$) eye movements. We recorded from 22 standard electrode sites spanning the scalp. We computed contralateral waveforms by averaging the activity recorded at right hemisphere electrode sites when subjects were cued to remember the left side of the memory array with the activity recorded from the left hemisphere electrode sites when they were cued to remember the right side. Contralateral delay activity was measured at posterior parietal, lateral occipital and posterior temporal electrode sites as the difference in mean amplitude between the ipsilateral and contralateral waveforms, with a measurement window of 300–900 ms after the onset of the memory array. Mean amplitudes were compared across conditions by analysis of variance.

Received 19 May; accepted 1 September 2005.

1. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
2. Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.* **74**, Whole No. 498 (1960).
3. Rainer, G., Asaad, W. F. & Miller, E. K. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* **393**, 577–579 (1998).
4. Bundesen, C., Pedersen, L. F. & Larsen, A. Measuring efficiency of selection from briefly exposed visual displays: a model for partial report. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 329–339 (1984).
5. Kane, M. J. & Engle, R. W. Working memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *J. Exp. Psychol. Gen.* **132**, 47–70 (2003).
6. Vogel, E. K. & Machizawa, M. G. Neural activity predicts individual differences in visual working memory capacity. *Nature* **428**, 748–751 (2004).
7. Vogel, E. K., Woodman, G. F. & Luck, S. J. Storage of features, conjunctions, and objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 92–114 (2001).
8. Cowan, N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* **24**, 87–185 (2001).
9. Pashler, H. Familiarity and visual change detection. *Percept. Psychophys.* **44**, 369–378 (1988).
10. Rypma, B. & D'Esposito, M. D. The influence of working memory demand and subject performance on prefrontal cortical activity. *J. Cogn. Neurosci.* **14**, 721–731 (2002).
11. Rypma, B. & D'Esposito, M. D. Isolating the neural mechanisms of age-related changes in human working memory. *Nature Neurosci.* **3**, 509–515 (2000).
12. Gale, A. & Edwards, J. in *Psychophysiology: Systems, Processes, and Applications* (eds Coles, M., Donchin, E. & Porges, S. W.) 431–486 (Guilford, New York, 1986).
13. Shih, S. & Sperling, G. Is there feature-based attentional selection in visual search? *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 758–779 (1996).
14. Anlo-Vento, L. & Hillyard, S. A. Selective attention to the colour and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Percept. Psychophys.* **58**, 191–206 (1996).
15. Woodman, G. F. & Vogel, E. K. Fractionating visual working memory: encoding and maintenance are independent processes. *Psychol. Sci.* **16**, 106–113 (2005).
16. Jiang, Y. & Kumar, A. Visual short-term memory for two sequential arrays: one representation or two representations? *Psychonomic Bull. Rev.* **11**, 495–500 (2004).
17. Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
18. Everling, S., Tinsley, C. J., Gaffan, D. & Duncan, J. Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neurosci.* **5**, 671–676 (2002).

19. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
20. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
21. Duncan, J. *et al.* A neural basis for general intelligence. *Science* **289**, 457–460 (2000).
22. Gray, J. R., Chabris, C. & Braver, T. S. Neural mechanisms of general fluid intelligence. *Nature Neurosci.* **6**, 316–322 (2003).
23. Engle, R. W., Kane, M. J. & Tuholski, S. W. in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (eds Miyake, A. & Shah, P.) 102–134 (Cambridge Univ. Press, New York, 1999).
24. Vogel, E. K., Luck, S. J. & Shapiro, K. L. Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 1656–1674 (1998).

Acknowledgements This work was supported by grants from the US National Institute of Mental Health and the Oregon Medical Research Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.K.V. (vogel@uoregon.edu).

LETTERS

Design principles of a bacterial signalling network

Markus Kollmann¹, Linda Løvdok², Kilian Bartholomé¹, Jens Timmer^{1,3} & Victor Sourjik²

Cellular biochemical networks have to function in a noisy environment using imperfect components. In particular, networks involved in gene regulation or signal transduction allow only for small output tolerances, and the underlying network structures can be expected to have undergone evolution for inherent robustness against perturbations¹. Here we combine theoretical and experimental analyses to investigate an optimal design for the signalling network of bacterial chemotaxis, one of the most thoroughly studied signalling networks in biology. We experimentally determine the extent of intercellular variations in the expression levels of chemotaxis proteins and use computer simulations to quantify the robustness of several hypothetical chemotaxis pathway topologies to such gene expression noise. We demonstrate that among these topologies the experimentally established chemotaxis network of *Escherichia coli* has the smallest sufficiently robust network structure, allowing accurate chemotactic response for almost all individuals within a population. Our results suggest that this pathway has evolved to show an optimal chemotactic performance while minimizing the cost of resources associated with high levels of protein expression. Moreover, the underlying topological design principles compensating for intercellular variations seem to be highly conserved among bacterial chemosensory systems².

Errors in signal transduction can lead to wrong development and behavioural decisions, and result in growth impairment or, in multicellular organisms, cancer¹. Therefore one expects a strong selective pressure towards networks showing intrinsic robustness against the various sources of inter- and intracellular perturbations, such as fluctuations in protein concentration. Owing to limitations in our quantitative understanding of most signalling networks, only few detailed studies of network robustness are currently available^{3–6}. In the simple signalling system of bacterial chemotaxis, activity of receptors on the cell surface is regulated by changes in ambient ligand concentration^{7,8}. Active receptors enhance autophosphorylation activity of the receptor-associated kinase CheA, which transmits the signal to the flagellar motors by phosphorylating a diffusible response regulator protein CheY. One of the key features of the chemotaxis pathway is precise adaptation—its ability to return to the same level of pathway activity under conditions of continuous stimulation. Failure of this systems property can lead to permanent swimming or tumbling behaviour and thus would restrict chemotactic response to a narrow range of chemoeffector concentrations. The most simple topology of a chemotactic signalling network that allows for precise adaptation is a two-state model proposed in ref. 3 (BL model). This model is schematically drawn in Fig. 1a. The core property of the model is that receptors can be reversibly methylated. A higher level of methylation increases the probability of a receptor to switch to an active state and thus balances the inhibiting effect of attractants. Almost perfect adaptation results from a constantly working methyltransferase (CheR) balanced by a methylesterase (CheB), which works only on active receptors⁹.

The BL model can readily explain the robustness of precise adaptation against a wide range of variations in kinetic parameters and protein concentrations as a consequence of integral feedback control at the methylation level^{3,9}. However, precise adaptation would be only physiologically relevant when the adapted level of CheY-P falls within the working range of the flagellar motor¹⁰. It is thus the stationary level of CheY-P rather than precise adaptation per se that should be under evolutionary selection, but in the BL model this quantity is not robust to large perturbations⁴.

By extending the BL model, it is possible to construct larger adaptive topologies having equal input–output characteristics (Fig. 1b–d) and analyse their robustness against intercellular variations in protein concentrations arising from protein synthesis. The latter represents the dominating source of variations in a bacterial cell population^{11,30}. Such intercellular variations can persist on the generation timescale because there is no evidence for an active degradation of chemotaxis proteins under standard growth and assay conditions, and the decrease in protein levels mainly results from dilution during cell division¹².

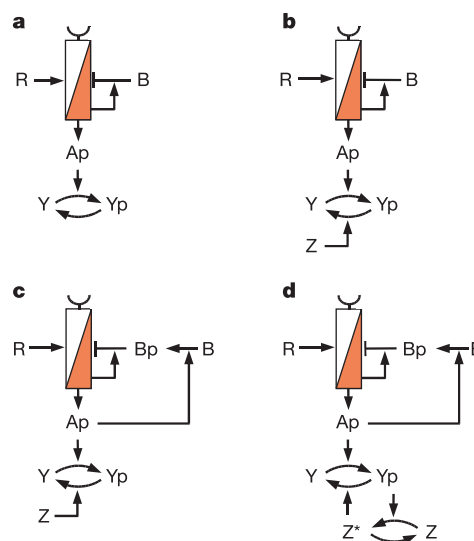


Figure 1 | Four possible network topologies of bacterial chemotaxis showing precise adaptation. Links between proteins indicate activations (arrows) or repressions (bar ends). The receptor can either be in an active state (red) or an inactive state (white). The proteins involved are denoted, for example, by A = CheA, and their phosphorylated forms by Ap = CheA-P. **a**, Minimal model as proposed in ref. 3. **b**, Same as model **a** but with a phosphatase CheZ substituting auto-dephosphorylation of CheY-P in topology **a**. **c**, Same as model **b** but only the phosphorylated form of CheB can form a complex with active receptors. This topology represents the experimentally established network of *E. coli*. **d**, Same as topology **c** but with an active form (Z^*) of the CheY phosphatase^{27,28}.

¹Institut für Physik, Universität Freiburg, Hermann-Herder-Str. 3, D-79104 Freiburg, Germany. ²ZMBH, University of Heidelberg, Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany. ³FDM, Universität Freiburg, Eckerstr. 1, D-79104 Freiburg, Germany.

To measure experimentally the amplitude of such intercellular variations in the levels of chemotaxis proteins, we replaced a native *cheY* gene in *E. coli* with *cheY* fused to a yellow fluorescent protein (*cheY-eyfp*) as a translational reporter. The reporter construct is thus expressed from the native promoter as part of polycistronic messenger RNA of the *meche* operon that also encodes Tar and Tap chemoreceptors, and most cytosolic chemotaxis proteins (Supplementary Fig. S1). CheA and CheW are expressed as part of another (*mocha*) operon. The protein levels in the population (Fig. 2a) show large intercellular variation with an apparently asymmetric distribution, consistent with previous observations¹². The gene expression noise $\eta = \sigma/n$, where σ is the standard deviation and n is the mean expression, depends on the level of gene transcription, and decreases from 0.67 to 0.47 upon deletion of the upstream transcription inhibitor, the anti-sigma factor FlgM^{13,14} (Fig. 2a). Recently, model-based analysis showed that transcription should dominate gene expression noise between proteins when expressed from the same mRNA transcript^{15,16}. We confirmed this conclusion by comparing the variation in the levels of CheY and CheZ expressed as fusions to yellow and cyan fluorescent proteins (YFP and CFP) from the same promoter in the same order as they are positioned on the chromosome and under their native ribosome-binding sequences (Fig. 2b). There is a strong co-variation in the expression levels of both, which arises from fluctuations in transcriptional activity, and only a moderate independent variation as a consequence of stochastic effects in both transcription and translation^{11,17}. Proteins expressed from two different chemotaxis operons, *mocha* and *meche*, which belong to the same regulon, also show significant co-variation

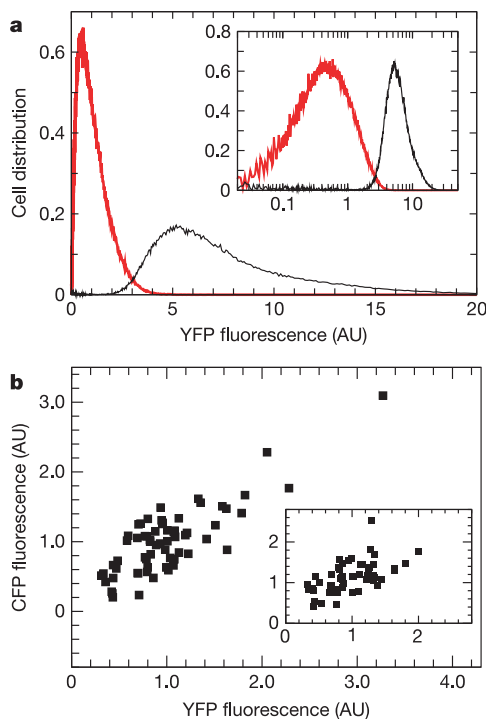


Figure 2 | Gene expression noise of the chemotaxis proteins. Intercellular variations in the level of CheY, expressed as a YFP fusion from the native chromosomal position. The distribution of wild-type (red line) and *flgM* (black line) cells is characterized by the means 1 and 6.6 and standard deviations 0.67 and 3.17, respectively. Inset: same data normalized to the same maximum intensity and displayed on a logarithmic scale. **b**, Correlation of the expression of CheY–YFP and CheZ–CFP from a single pTrc promoter without induction. Values for uncorrelated and concerted variations of the gene expression noise are $\eta_{in} = 0.20$ and $\eta_{ex} = 0.44$, respectively. Inset: correlation of the expression of CheY–YFP and CheA–CFP from the native chromosomal positions, with $\eta_{in} = 0.26$ and $\eta_{ex} = 0.35$. AU, arbitrary units.

in gene expression noise (Fig. 2b inset) but with a 25% increase of independent variations in comparison with proteins expressed from the same operon.

We further experimentally determined the effect of co-variation in the levels of all signalling proteins on chemotactic behaviour (Fig. 3). Concerted overexpression of all proteins up to 6.6-fold above the native level had little effect on chemotaxis efficiency, as measured by a chemotaxis-driven spreading of bacteria in an attractant gradient created by nutrient depletion in soft agar (swarm assay). Thus, the CheY-P concentration in the overexpressing cells must be in the working range of the flagellar motor. This conclusion was further confirmed by the observation that the average time the motor spends rotating clockwise (clockwise bias) was nearly unaffected by protein overexpression (Fig. 3b). Notably, the corresponding standard deviation decreased with the level of expression, as expected from the decline in strength of gene expression noise (Fig. 2 and Supplementary Fig. S3).

Relying on these experimental data, we analysed mathematically the robustness of the four network topologies drawn in Fig. 1 under conditions of physiological perturbations. Owing to a steep response of the flagellar motor¹⁰, the level of CheY-P in adapted cells can vary only about one-third from its optimal value. Outside this concentration regime the cell mainly tumbles or swims continuously and cannot properly respond to stimuli. For example, a fully tumbling cell cannot respond to repellents and has a reduced sensitivity to attractants. The chemotactic efficiency of a population therefore depends on the fraction of bacteria for which the CheY-P levels are within these limits. We tested the different hypothetical network topologies, shown in Fig. 1, for their ability to reproduce qualitatively the experimental data for concerted variations of mean expression levels (Fig. 3). In the computer simulations, we used the experimentally determined gene expression noise (Fig. 2) as an estimate for the minimum intercellular variations. The established topology for *E. coli* (Fig. 1c) reproduces accurately the experimental data, whereas the simpler topology (Fig. 1a) fails to match the data. To quantify further the robustness of the different network topologies, we calculated the fraction of fully chemotactic bacteria for different strengths of gene expression noise (Fig. 4). The topologies in Fig. 1c, d allow for

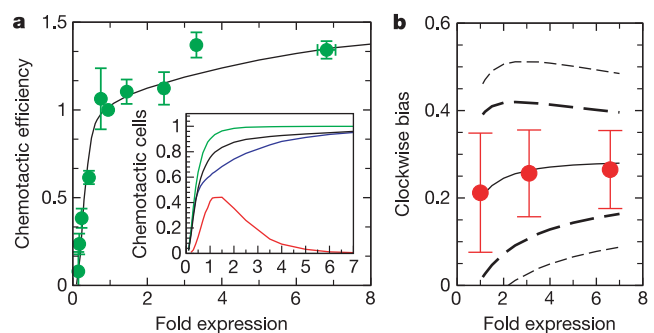


Figure 3 | Effect of the total concentration of signalling proteins on chemotaxis. **a**, Chemotaxis efficiency of *flgM* (VS102) cells, expressing varying levels of FlgM from a plasmid, determined in a swarm assay and normalized to the value of wild-type (RP437) cells. The relative mean expression of chemotaxis proteins at each FlgM level was measured as in Fig. 2a, using the LL1 strain as a reporter. The solid line is a guide to the eye, and error bars indicate standard errors. Inset: simulated fraction of fully chemotactic cells for topologies in Fig. 1a (red line), b (blue line), c (black line) and d (green line). **b**, Clockwise motor bias as a function of expression of chemotaxis proteins. Each point represents a mean of 20–30 cells. Error bars indicate standard deviations and illustrate intercellular variation. The solid line shows the clockwise bias calculated from the mathematical model of Fig. 1c under gene expression noise estimated from Fig. 2. The calculated standard deviation is depicted for the topology in Fig. 1c, assuming a steepness of the motor response curve with a Hill coefficient of five²⁹ (thick dashed lines) or ten¹⁰ (thin dashed lines).

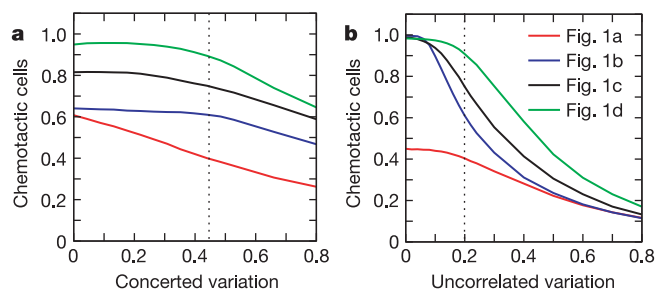


Figure 4 | Simulated fraction of chemotactic cells as a function of gene expression noise. Topologies are marked as in Fig. 3. The native level of gene expression noise is indicated by dotted lines. **a**, Concerted variations ranging up to twofold of the wild-type strength with uncorrelated variations kept at the wild-type value, $\eta_{in} = 0.2$. **b**, Same as **a** but with varying strength of uncorrelated variations and with concerted variations fixed to the wild-type value, $\eta_{ex} = 0.44$.

the highest fraction of cells in the population to respond accurately to changes in ligand concentration. Furthermore, the topologies in Fig. 1b–d are sufficiently robust to compensate for co-variations in expression levels (Fig. 4a) but tolerate only a moderate increase in independent fluctuations in protein concentrations (Fig. 4b).

There are two key features accounting for the higher robustness of the topologies shown in Fig. 1c, d. First, robustness against concerted variations requires a balance of phosphatase and kinase reactions and similar requirements for the methylation process, as shown by mathematical analysis in Box 1. The condition for a robust adaptive chemotaxis pathway is that CheY-P demands a phosphatase (CheZ) whereas CheB-P must not have one. But this kind of robustness is only valid for concentrations larger than the wild-type level (Fig. 3a). The strong decrease in the number of chemotactic bacteria at lower expression levels (Fig. 3a inset) arises because the total CheY concentration drops below the working range of the flagellar motor. The increase in chemotactic performance of the topologies in Fig. 1b–d with the mean expression level can be explained by an accompanying decline of the gene expression noise¹¹ (Fig. 2a and

Supplementary Fig. S3). The location of the wild-type concentration as shown in Fig. 3a seems to reflect a selective pressure towards overall low protein concentrations¹⁸.

The second topological feature leading to higher swarming efficiency is CheB phosphorylation resulting in an additional negative feedback loop (Fig. 1c, d). As shown in Box 2, this second feedback loop compensates partially for deviations from the optimal CheY-P level without changing the input–output characteristics. CheB phosphorylation has been previously shown to be non-essential for adaptation^{3,4}, and our analysis suggests that its main function might be in noise reduction. The essential features for the two design principles described above seem to be present among most established pathway topologies of bacterial chemotaxis². In particular, the CheB phosphorylation feedback is almost universal, and although many bacteria lack CheZ, the function of CheY phosphatase is taken over by another protein or by the kinase itself. The even higher robustness of the topology in Fig. 1d arises from the activation of CheZ by CheY-P that follows the same noise compensatory mechanism as the CheB phosphorylation feedback.

Box 2 | Error reduction mechanisms

Errors in the output signal arising from uncorrelated variations of protein levels and deviations from the optimal rate constants can be partially compensated by additional negative feedback loops²² (Fig. 1). Here, we focus on the gain in robustness due to activation of the methyltransferase (CheB) by phosphotransfer from CheA. We simplify the *E. coli* chemotaxis topology (Fig. 1c) such that any receptor has only one methylation site and the average activity of methylated receptors depends on the ambient chemoeffector concentration. Non-methylated receptors remain inactive. Using Michaelis-Menten kinetics, the steady-state equation for the methylation process reads

$$\partial_t T_M = k_R R - k_B B p \frac{T_A}{K_B + T_A} = 0 \quad (3)$$

with k_R and k_B the associated rate constants and K_B the Michaelis-Menten constant for binding of CheB-P to active receptor complexes³. The concentrations of methylated and active receptors are denoted by T_M and T_A , and the concentrations of the methyltransferase and active methyltransferase by R and Bp , respectively. For the topologies in Fig. 1c, d the methyltransferase, CheB, is active only in the phosphorylated form and thus depends on the concentration of the phosphodonor, $Ap = [\text{CheA-P}]$. We have also assumed that the methyltransferase, CheR, works at saturation. The steady-state equation for Ap is given by

$$k_A T_A (A^T - Ap) - k_Y Ap (Y^T - Yp) = 0 \quad (4)$$

with $Yp = k_Y Ap Y^T / (k_Y Ap + k_Z Z)$ and where $k_Z Z$ is the dephosphorylation rate of CheY-P. The superscript T indicates total protein concentrations. In the above equation we neglected the small contribution of the phosphoacceptor CheB as (ref. 23) $k_Y Y^T \gg k'_B B^T$, with k_Y and k'_B the rates of phosphate transfer from CheA-P to CheY and CheB, respectively. As Ap and R are linked through equations (3) and (4), a small increase in the amount of methyltransferase, $R + \Delta R$, results in a change of phosphorylated receptors given by

$$\Delta Ap = \left[\alpha + \beta \frac{\partial Bp}{\partial Ap} \right]^{-1} \gamma \Delta R \quad (5)$$

To arrive at equation (5) we have performed a linear expansion around fixed values for the remaining protein concentrations. The linear expansion coefficients α and β can be shown to have equal sign (see Supplementary Information). The derivative $\partial Bp / \partial Ap > 0$ manifests the higher robustness against perturbations of the network topologies in Fig. 1c, d. This contribution increases the amount of active methyltransferase (CheB-P) whenever the average activity of the receptors is rising, and thus partially compensates for an increased amount of methyltransferase. This term is absent in the topologies of Fig. 1a, b, as here the methyltransferase is not phosphorylated and thus $\partial Bp / \partial Ap = 0$. Fluctuations in other protein levels are compensated for in an equivalent way.

Box 1 | Robustness against variations in transcriptional activity

A general deterministic description for the concentrations of the N different phosphorylation and methylation states $y(t) = \{y_1(t), \dots, y_N(t)\}$ of a signalling pathway are given by the equations

$$\partial_t y_i(t) = F_i(y(t)|x^T) \quad (1)$$

(see Supplementary Information for details and definitions). The sum over different states, $(y_i)_k$, of the protein with index k is connected to its total concentration by $\sum_{\{y_i\}_k} y_i(t) = x_k^T$. For the stationary solution, $F_i(y(t)|x^T) = 0$, to be invariant against co-varying total protein concentrations, for example due to a λ -fold change in transcriptional activity $x^T = (x_1^T, \dots, x_M^T) \rightarrow (\lambda x_1^T, \dots, \lambda x_M^T)$ of the M chemotaxis proteins, we have to demand homogeneity of F with respect to x^T ,

$$F_i(y(t)|\lambda x^T) = \lambda^{\mu_i} F_i(y(t)|x^T) \quad (2)$$

with $\mu_i > 0$. For the equations corresponding to the topology in Fig. 1a the homogeneity condition can not be satisfied. For the topologies in Fig. 1b–d we have the case $F_i(y(t)|\lambda x^T) \approx \lambda F_i(y(t)|x^T)$ for $\lambda > 1$ and therefore these topologies are invariant against changes in transcriptional activity of signalling proteins at expression levels higher than the wild type. For example, to keep the level of CheY-P invariant under a λ -fold overexpression, the resulting increase of phosphotransfer to CheY has to be balanced by an increase in dephosphorylation rate. This is achieved by a simultaneous upregulation of the phosphatase CheZ. The conditions of invariance and the consequences of their violation are discussed in the Supplementary Information.

Reflecting strong selection, the chemotaxis pathway in *E. coli* seems to be optimized for high sensitivity, fast response and perfect adaptation^{7,8,19,20}. As shown in this work, the experimentally established design of the chemotaxis network in *E. coli* represents a minimal topology providing high robustness to physiological perturbations. This network design can compensate for the observed strong co-variations in gene expression but the negative effect of uncorrelated variations on the efficiency of chemotaxis can only be attenuated. Similar correlations in expression have been found recently in eukaryotes for genes under identical control²¹. We can therefore expect that analogous design principles, compensating for intercellular variations, will apply to all signalling networks and gene regulation systems whenever precise regulation of an output signal is demanded.

METHODS

Bacterial strains and plasmids. All strains used in this study were derived from a wild-type chemotaxis strain RP437 using a pAMPTs homologous recombination system of allele exchange, described before²⁴. Strain VS162 carries a *cheY-eyfp* fusion construct in place of a *cheY* gene on the chromosome, and strain LL6 carries *cheY-eyfp* and *cheA-ecfp* fusion constructs in place of *cheY* and *cheA* genes on the chromosome. Strains VS102 (*flgM*) and LLI (*cheY-eyfp flgM*) carry in-frame deletions of an anti-sigma factor (*FlgM*) that negatively controls the expression of all chemotaxis and flagellar (class III) genes. Plasmid pLL16 Amp^r encodes *FlgM* expressed under control of an isopropyl β -D-thiogalactoside (IPTG)-inducible promoter pTrc. Plasmid pVS88 encodes CheY-YFP and CheZ-CFP fusion proteins transcribed as one polycistronic mRNA from the pTrc promoter²⁵. The level of CheY-YFP expression in the absence of IPTG closely matched the expression from a native promoter (VS162).

Growth conditions. All strains were grown under standard chemotaxis conditions at 34 °C in tryptone broth (TB) as described before^{24,25} in the presence of varying amounts of IPTG. Swarm assays were performed at 34 °C on TB plates supplemented with 0.3% agar (Applichem) and indicated concentrations of IPTG.

Quantification of gene expression. Expression of fluorescent reporter proteins in individual cells was quantified using flow cytometry on a FACScan (BD Biosciences) equipped with a 488-nm argon laser, or fluorescence imaging on an Axiovert 200 fluorescence microscope equipped with an ORCA AG CCD camera (Hamamatsu). FACScan data were analysed using CellQuest Pro 4.0.1 software. Imaging data were analysed using ImageJ software (W. Rasband) to quantify fluorescence of the entire cell. When tested on the same population, both methods gave essentially identical results.

Tethering. Cells were tethered to a glass coverslip as described earlier²⁶. The rotation of tethered cells was recorded for 3 min at 25 °C with a microscope (Zeiss Axiostar plus) equipped with a CCD camera (Panasonic WV-BP330) and a DV Recorder (Panasonic AG-DV1DC). The analysis was performed using IPS image recognition software from the Visometrics Group.

Description of the mathematical model. The dynamical behaviour of the networks was described on the level of ordinary differential equations assuming Michaelis-Menten kinetics. The kinetic constants were taken from *in vivo* and *in vitro* measurements, except for the methylation process and the CheY-CheZ feedback loop shown in Fig. 1d, where constants were determined from optimization of a population of 70 cells with respect to their least deviation from the midpoint of the motor response curve¹⁰ (3.2 μ M) under the experimentally determined gene expression noise. Fully chemotactic cells were defined as cells for which the level of CheY-P was within the interval 2.2–4.3 μ M and for determination of their proportion in Figs 3 and 4 a population size of 10⁴ cells was assumed. A detailed description of the mathematical model is given in the Supplementary Information.

Received 25 May; accepted 12 September 2005.

1. Kitano, H. Biological robustness. *Nature Rev. Genet.* **5**, 826–837 (2004).
2. Szurmant, H. & Ordal, G. W. Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol. Mol. Biol. Rev.* **68**, 301–319 (2004).
3. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).

4. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).
5. von Dassow, G., Meir, E., Munro, E. M. & Ordell, G. M. The segment polarity network is a robust developmental module. *Nature* **406**, 188–191 (2000).
6. Eldar, A. *et al.* Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. *Nature* **419**, 304–308 (2002).
7. Sourjik, V. Receptor clustering and signal processing in *E. coli* chemotaxis. *Trends Microbiol.* **12**, 569–576 (2004).
8. Wadhams, G. H. & Armitage, J. P. Making sense of it all: bacterial chemotaxis. *Nature Rev. Mol. Cell Biol.* **5**, 1024–1037 (2004).
9. Mello, B. & Tu, Y. Perfect and near-perfect adaption in a model of bacterial chemotaxis. *Biophys. J.* **84**, 2943–2956 (2003).
10. Cluzel, P., Surette, M. & Leibler, S. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science* **287**, 1652–1655 (2000).
11. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Nature* **297**, 1183–1187 (2002).
12. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. & Elowitz, M. B. Gene regulation at the single-cell level. *Nature* **307**, 1962–1965 (2005).
13. Kalir, S. *et al.* Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292**, 2080–2083 (2001).
14. Aldridge, P. & Hughes, K. T. Regulation of flagellar assembly. *Curr. Opin. Microbiol.* **5**, 160–165 (2002).
15. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
16. Swain, P. S. Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol.* **344**, 965–976 (2004).
17. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
18. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592 (2005).
19. Bray, D., Levin, M. D. & Morton-Firth, C. J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85–88 (1998).
20. Webre, D. J., Wolanin, P. M. & Stock, J. B. Bacterial chemotaxis. *Curr. Biol.* **13**, 47–49 (2003).
21. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
22. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
23. Sourjik, V. & Berg, H. C. Receptor sensitivity in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **99**, 123–127 (2002).
24. Sourjik, V. & Berg, H. C. Localization of components of the chemotaxis machinery of *Escherichia coli* using fluorescent protein fusions. *Mol. Microbiol.* **37**, 740–751 (2000).
25. Sourjik, V. & Berg, H. C. Functional interactions between receptors in bacterial chemotaxis. *Nature* **428**, 437–441 (2004).
26. Turner, L., Caplan, S. R. & Berg, H. C. Temperature-induced switching of the bacterial flagellar motor. *Biophys. J.* **71**, 2227–2233 (1996).
27. Blat, Y., Gillespie, G., Bren, A., Dahlquist, F. W. & Eisenbach, M. Regulation of phosphatase activity in bacterial chemotaxis. *J. Mol. Biol.* **284**, 1191–1199 (1998).
28. Almog, G., Stone, L. & Ben-Tal, N. B. Multi-stage regulation, a key to reliable adaptive biochemical pathways. *Biophys. J.* **81**, 3016–3028 (2001).
29. Scharf, B. E., Fahrner, K. A., Turner, L. & Berg, H. C. Control of direction of flagellar rotation in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **95**, 201–206 (1998).
30. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Shimizu, M. D. Levin, K. Lipkow and F. Geier for comments on the manuscript, and C. Bechinger, S. Bleil and S. Schulmeister for technical help. This work was supported by ZMBH funding, DFG grants to V.S. and K.B., and the BMBF project 'Systems of Life-Systems Biology'.

Author Contributions V.S. designed the experiments and L.L. carried them out. M.K. performed the mathematical modelling together with K.B. The paper was written by M.K. and V.S. with comments from J.T. L.L. and K.B. contributed equally to this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.K. (markus.kollmann@fdm.uni-freiburg.de) or V.S. (v.sourjik@zmbh.uni-heidelberg.de).

LETTERS

The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation

Uttiya Basu^{1*}, Jayanta Chaudhuri^{1*†}, Craig Alpert¹, Shilpee Dutt², Sheila Ranganath¹, Gang Li¹, Jason Patrick Schrum¹, John P. Manis² & Frederick W. Alt¹

Antibodies, which are produced by B-lineage cells, consist of immunoglobulin heavy (IgH) and light (IgL) chains that have amino-terminal variable regions and carboxy-terminal constant regions. In response to antigens, B cells undergo two types of genomic alterations to increase antibody diversity. Affinity for antigen can be increased by introduction of point mutations into IgH and IgL variable regions by somatic hypermutation. In addition, antibody effector functions can be altered by changing the expressed IgH constant region exons through IgH class switch recombination (CSR)^{1–3}. Somatic hypermutation and CSR both require the B-cell-specific activation-induced cytidine deaminase protein (AID)^{4–6}, which initiates these reactions through its single-stranded (ss)DNA-specific cytidine deaminase activity^{7–11}. In biochemical assays, replication protein A (RPA), a ssDNA-binding protein¹², associates with phosphorylated AID from activated B cells and enhances AID activity on transcribed double-stranded (ds)DNA containing somatic hypermutation or CSR target sequences. This AID–RPA association, which requires phosphorylation, may provide a mechanism for allowing AID to access dsDNA targets in activated B cells^{13,14}. Here we show that AID from B cells is phosphorylated on a consensus protein kinase A (PKA) site and that PKA is the physiological AID kinase. Thus, AID from non-lymphoid cells can be functionally phosphorylated by recombinant PKA to allow interaction with RPA and promote deamination of transcribed dsDNA substrates. Moreover, mutation of the major PKA phosphorylation site of AID preserves ssDNA deamination activity, but markedly reduces RPA-dependent dsDNA deamination activity and severely impairs the ability of AID to effect CSR *in vivo*. We conclude that PKA has a critical role in post-translational regulation of AID activity in B cells.

AID from activated splenic B cells (AID^{Bcell}) is phosphorylated and interacts with RPA, whereas AID generated by overexpression in non-lymphoid 293 cells (AID²⁹³) is not highly phosphorylated and does not functionally interact with RPA¹³. Thus, B cells seem to express a kinase that can phosphorylate AID and, thereby, modulate its activity. To elucidate the AID kinase, we incubated AID²⁹³ with nuclear extracts from either 293 cells or CSR-activated AID-deficient B cells in the presence of [γ -³²P]ATP. The activated B cell, but not 293, extracts contained an activity that phosphorylated AID, as evidenced by the presence of a radiolabelled polypeptide that migrated with the same mobility as AID on SDS gels (Fig. 1a). Notably, *in vitro* phosphorylated AID²⁹³ associated with RPA present in 293 extracts, and this association was blocked by phosphatase treatment (Fig. 1b). Therefore, activated B-cell extracts contain an AID kinase that imparts upon AID²⁹³ the ability to interact with RPA. Although

AID^{Bcell} is more abundant in the cytoplasm than in the nucleus, as judged by ssDNA deaminase activity (Fig. 1c) and western blot analyses (Supplementary Fig. 3), nuclear AID had a significantly higher dsDNA deamination activity (Fig. 1c). Because dsDNA deamination activity correlates with the phosphorylation of AID¹³, the nuclear AID pool seems to be enriched for the functionally phosphorylated form of AID. Therefore, to identify sites phosphorylated on AID^{Bcell}, we performed a mass spectrometric analysis on purified nuclear AID from activated B cells.

The mass spectrometric analysis revealed that nuclear AID^{Bcell} is phosphorylated on serine 38 (S38) and tyrosine 184 (Y184) (Fig. 1d; see also Supplementary Methods for details). In contrast, we failed to detect any phosphorylated residue on AID²⁹³ (data not shown). Notably, the S38 residue is located within a cyclic AMP-dependent PKA consensus phosphorylation site, RRX(S/T), where R is an invariant arginine, X is any amino acid and S/T a serine or threonine¹⁵. This site is conserved in AID across all species that undergo CSR (Supplementary Fig. 1a). In addition, all species that express AID have an additional potential PKA site at T27, which is a classical site in human, chimpanzee and chicken, but varies from the consensus in other species such as mouse (for example, RHET; Supplementary Fig. 1a). To address further the functionality of these sites, we assayed peptides representing mouse AID S38 and T27 sites for the ability to be phosphorylated by recombinant PKA, and found that both were PKA substrates (Supplementary Fig. 1b). We conclude that S38 is a functional PKA site on murine AID in B cells activated for CSR, and that T27 has the potential to serve as a PKA site.

To investigate further the relationship between AID and PKA, we subjected B-cell extracts to glycerol gradient sedimentation and assayed gradient fractions for the ability to promote the interaction of AID²⁹³ with RPA. Fraction 19 contained AID-activating activity along with maximum PKA catalytic subunit (PKA_{cat}) protein, as determined by western blot analysis (Fig. 2a, top and bottom). Furthermore, AID kinase activity co-eluted with PKA_{cat} from a CM-sepharose column (data not shown). Finally, immunoprecipitates of AID from total B-cell extracts contained an activity capable of phosphorylating a consensus PKA peptide substrate (Fig. 2b, left panel, top), and also contained the PKA regulatory subunit (Fig. 2b, right panel, top). Thus, both co-purification and co-immunoprecipitation experiments indicate that AID and PKA functionally associate in activated B cells.

In vitro assays demonstrated that purified PKA_{cat} could directly phosphorylate purified AID²⁹³ *in vitro* (Fig. 2c). Moreover, such phosphorylation also allowed AID²⁹³ to acquire dsDNA deamination activity in the presence of RPA. Thus, T7 polymerase-aided transcription of a synthetic RGYW somatic hypermutation substrate in

¹The Howard Hughes Medical Institute, The Children's Hospital, The CBR Institute for Biomedical Research, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Joint Program in Transfusion Medicine, Children's Hospital, Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. [†]Present address: Immunology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA.

*These authors contributed equally to this work.

the presence of AID²⁹³ and RPA yielded little AID-mediated dsDNA deamination; whereas pre-phosphorylation of AID²⁹³ by PKA *in vitro* resulted in a 25-fold increase in dsDNA deamination activity (Fig. 2d). As a control, treatment of phosphorylated AID²⁹³ with alkaline phosphatase abrogated dsDNA deamination activity (Supplementary Fig. 2b). As a further test of whether PKA is the AID kinase, we tested the sensitivity of B-cell nuclear extracts to a PKA-specific inhibitor (H89) and found that H89 pre-treatment blocked the ability to phosphorylate AID²⁹³ (Fig. 1a). Notably, treatment of CH12F3 cells with H89 severely abrogated the interleukin-4 (IL-4)-, transforming growth factor (TGF)- β - and anti-CD40-induced CSR to IgA¹⁶, even though AID expression and germline transcripts for the IgH α constant region (C α) occurred at levels similar to those of untreated cells (Supplementary Fig. 4). Together, the latter results are consistent with the interpretation that H89 inhibits CSR in CH12F3 cells by inhibiting PKA-mediated AID phosphorylation, although it

is not possible to firmly rule out an inhibitory role of potential nonspecific factors, such as effects on proliferation¹⁷.

To test directly whether the S38 and Y184 residues, as well as the T27 residue, are important for the interaction with RPA and for the dsDNA deamination activity, we assayed partially purified AID²⁹³ proteins in which we individually mutated the T27, S38 and Y184 residues to alanine. When expressed at similar levels (Fig. 3a, b, bottom panels), the mutants had substantial catalytic ssDNA deamination activity, although the T27A mutant was less active (Fig. 3d). However, both the T27A and the S38A mutants failed to undergo PKA phosphorylation, whereas the Y184A mutant was phosphorylated at similar levels to wild-type AID (Fig. 3b). The effect of the T27A mutation on S38 phosphorylation may reflect site-interdependent phosphorylation processes^{18–20} or simply a nonspecific structural alteration in AID that inhibits S38 phosphorylation. As

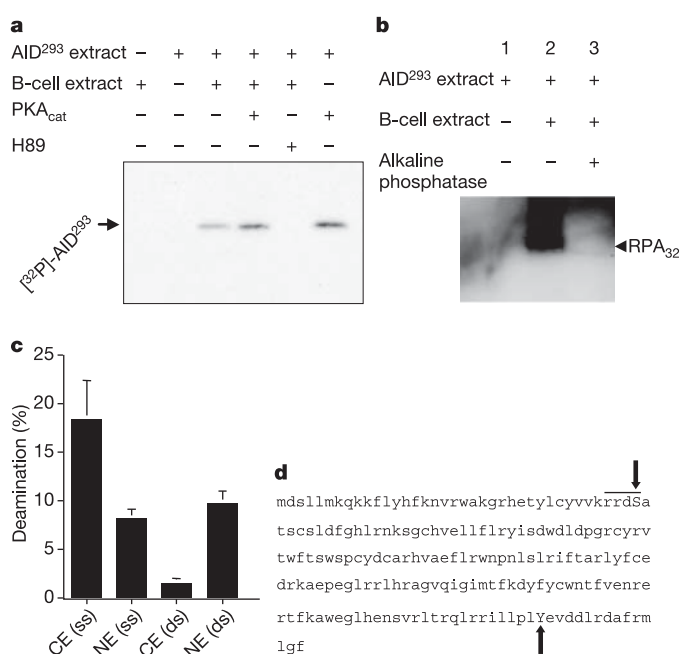


Figure 1 | *In vitro* phosphorylation of AID²⁹³ and determination of AID phosphorylation sites. **a**, AID²⁹³ is phosphorylated by activated B-cell nuclear extract. 293 cells were either mock-transfected (–) or transfected with AID cDNA and expressed from a CMV promoter. The cell extracts were incubated with nuclear extracts from 293 cells or from AID-deficient splenic B cells in the presence of [γ -³²P]ATP. Where indicated, the reaction mixtures also contained the catalytic subunit of PKA (PKAc) and the PKA inhibitor H89. AID²⁹³ was immunoprecipitated from the phosphorylated protein extracts using anti-AID antibodies and then analysed by SDS–polyacrylamide gel electrophoresis followed by autoradiography of the dried gels. **b**, Phosphorylated AID²⁹³ interacts with RPA. Immunoprecipitated AID²⁹³ was phosphorylated *in vitro* with 293 cell nuclear extract (lane 1) or B-cell nuclear protein fraction (lane 2) and then incubated with 293 whole-cell extracts. Proteins that bound to the phosphorylated AID immunoprecipitate were analysed by western blotting using antibodies against the 32-kDa subunit of RPA. This interaction is sensitive to calf alkaline phosphatase treatment (lane 3). **c**, Nuclear extracts of B cells have significantly more dsDNA deamination⁷ activity than cytoplasmic extracts. Twenty micrograms of partially purified nuclear extract (NE) or cytoplasmic extract (CE) derived from activated B cells was assayed for the deamination of ssDNA or transcribed dsDNA in the presence of RPA. Values represent the mean from three experiments and error bars represent the standard deviation from the mean. **d**, AID purified from B cells was analysed by mass spectrometry. Identified phosphorylated residues (serine 38 and tyrosine 184) are indicated by arrows; the PKA consensus site is indicated with a line above it.

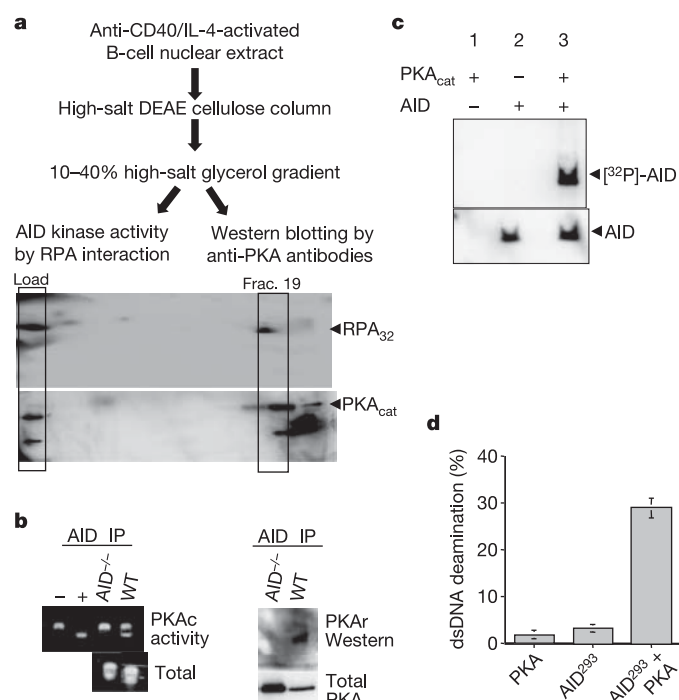


Figure 2 | AID is phosphorylated by PKA. **a**, Scheme of partial purification of AID kinase (top). Fractions containing AID kinase activity (upper panel, western blot with anti-RPA₃₂ antibodies in AID immunoprecipitates) also contain PKA (lower panel, western blot with anti-PKA_{cat} antibodies) when B-cell nuclear extract is sedimented through a 10–40% glycerol gradient. The box represents fraction 19 obtained from the glycerol gradient. **b**, AID interacts with PKA in B cells. AID–PKA complex was immunoprecipitated from wild-type or AID^{–/–} B-cell extracts using anti-AID antibodies. The immunoprecipitated (IP) complex was assayed for PKA activity by observing the mobility of a PKA substrate peptide on a 0.8% agarose gel after phosphorylation (left panel, top). The unphosphorylated peptide (–) and the phosphorylated peptide (+) migrate in opposite directions²⁸. The immunoprecipitate was analysed on a 15% SDS–PAGE followed by western blotting with anti-PKAr (regulatory) antibodies for detection of the regulatory subunit of PKA. **c**, The purified recombinant catalytic subunit of PKA can phosphorylate purified AID²⁹³ *in vitro*. AID²⁹³ (lanes 2, 3) or buffer alone (lane 1) was incubated with the catalytic subunit of PKA (lanes 1 and 3) in the presence of [γ -³²P]ATP; the reaction products were analysed by SDS–PAGE followed by transfer to an immobilon P membrane and autoradiography (top panel). The membrane was western blotted with anti-AID antibodies. **d**, AID²⁹³ phosphorylated by PKA has dsDNA deamination activity. AID²⁹³ alone, PKA alone or AID²⁹³ and PKA together were incubated with ATP and then assayed for dsDNA deaminase activity^{7,13}. Values represent the mean from three independent experiments and error bars represent the standard deviation from the mean.

expected, based on lack of phosphorylation, both the T27A and the S38A mutants, after incubation with PKA, failed to interact with the 32-kDa subunit of the trimeric complex RPA (RPA₃₂) in *in vitro* binding assays (Fig. 3a). Furthermore, when partially purified preparations of these mutants were assayed for RPA-dependent dsDNA deamination activity, only the Y184A mutant was active (Fig. 3c). Finally, we purified the wild-type and S38A mutant proteins to homogeneity, and confirmed that the purified S38A mutant protein exhibits the same properties as those observed in the partially purified extracts (Supplementary Fig. 2c–e). In particular, the kinetics of ssDNA deaminase activity for wild-type and S38A AID were comparable when assayed at different concentrations (Supplementary Fig. 2a).

Because phosphorylated AID^{Bcell} was purified from B cells that were activated to undergo CSR, we investigated the effects of mutating phosphorylation sites (T27A, S38A and Y184) on CSR in activated B cells. For this purpose, we stimulated AID-deficient splenic B cells with IL-4 and anti-CD40 to induce CSR to IgG1, followed by retroviral infection of wild-type or mutant AID together with green fluorescent protein (GFP) via an internal ribosome entry site (IRES) for detection of infected cells. Introduction of wild-type AID or AID(Y184A) mutant restored CSR in AID^{-/-} cells (Fig. 4a, b).

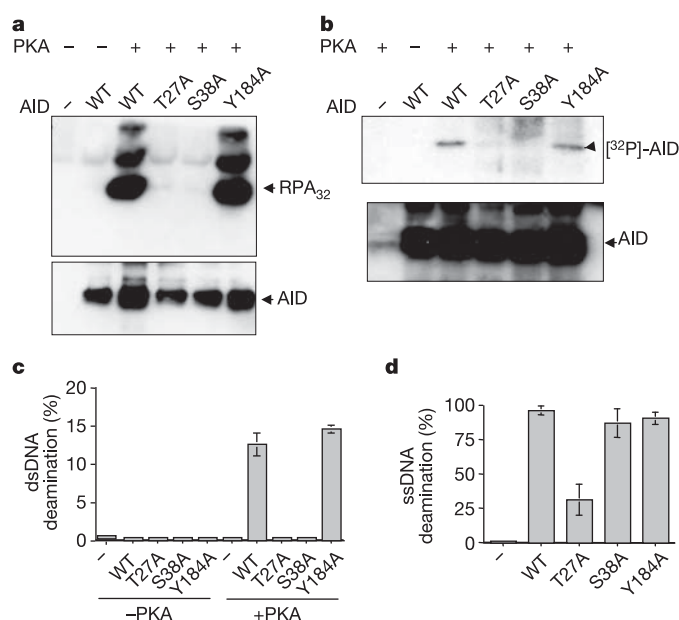


Figure 3 | AID phosphorylation mutants fail to interact with RPA. **a**, AID mutants expressed in 293 cells were phosphorylated by PKA and immunoprecipitated with anti-AID antibodies, then analysed for interaction with RPA as described in the legend to Fig. 1 (top). The presence of two minor bands along with the RPA₃₂ major band is due to the high concentration of RPA₃₂ in the immunoprecipitate. The amount of AID present in the extracts used in the immunoprecipitation was assessed by western blotting using anti-AID antibodies (bottom). **b**, 293 cell extracts expressing wild-type or mutant AID were phosphorylated by PKA in the presence of [γ -³²P]ATP followed by immunoprecipitation of the protein with anti-AID antibodies. The immunoprecipitate was resolved by SDS-PAGE, electrophoretically transferred to a membrane, followed by autoradiography (top) or western blotting (bottom) to detect the phosphorylated and immunoreactive band of AID, respectively. **c**, **d**, Phosphorylation mutants of AID (T27A or S38A or Y184A) and wild-type AID were expressed in 293 cells and then partially purified (through DEAE cellulose and glycerol gradient)^{7,13}. The proteins were assayed for RPA-dependent dsDNA deamination activity after phosphorylation by recombinant PKA (**c**), or ssDNA deamination activity without addition of PKA^{7,13} (**d**). For **c** and **d**, values represent the mean from three independent experiments and error bars represent the standard deviation from the mean.

In contrast, when expressed at similar levels (Fig. 4c), AID(T27A) and AID(S38A) did not detectably rescue CSR (Fig. 4a, b), indicating that integrity of the PKA phosphorylation site(s) is required for AID-mediated CSR *in vivo*.

We conclude that PKA phosphorylates the serine 38 residue on mouse AID and that this phosphorylation event activates AID to enable interaction with RPA, thereby augmenting the ability of AID to effect CSR *in vivo*. RPA can target AID to transcribed switch regions and may also be involved in pathways downstream of deamination^{13,21,22}. Although the requirement of AID phosphorylation for RPA interaction suggests a potential role in somatic hypermutation, this possibility can only be addressed by gene-targeted mutation of the S38 residue of AID *in vivo*. Notably, S38 is conserved in the AID protein of higher vertebrates that undergo

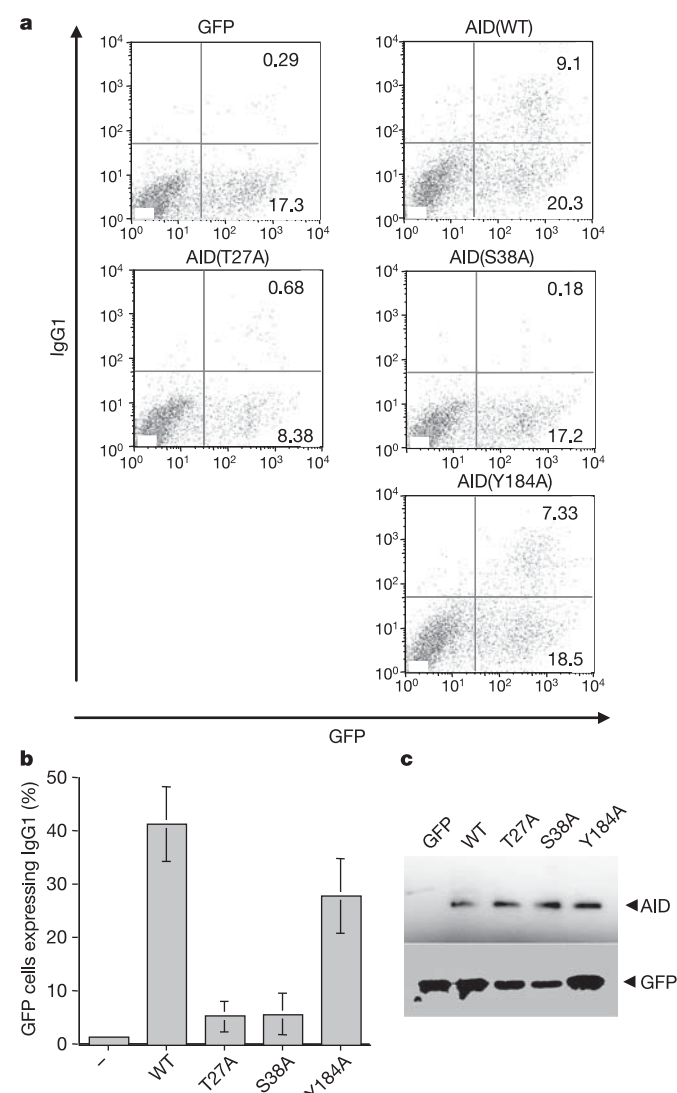


Figure 4 | Mutations of the AID phosphorylation site impair CSR *in vivo*. **a**, AID-deficient B cells, stimulated with anti-CD40 and IL-4, were infected with a retrovirus expressing wild-type AID, phosphorylation site mutant AID(T27A), AID(S38A) or AID(Y184A), or a retrovirus expressing GFP alone²⁹; the level of CSR to IgG1 was evaluated by a FACS analysis. **b**, The percentage of GFP-positive cells that had undergone CSR to IgG1 from four independent experiments was plotted to determine the activity of the mutant proteins. Error bars represent the standard deviation from the mean. **c**, Extracts were analysed from the retrovirally infected B cells by western blotting using anti-AID (top panel) or anti-GFP (bottom panel) antibodies to determine expression levels of the mutant proteins.

both somatic hypermutation and CSR, but not in that of lower vertebrates, such as bony fish, that only undergo somatic hypermutation (Supplementary Fig. 1a). Given that AID from fish can support some CSR in mice²³, the presence of an aspartate residue in fish at a similar location to S38 in higher vertebrates (Supplementary Fig. 1a) raises the possibility of constitutive mimicry of S38 phosphorylation. Although transcription of AID is tightly regulated, additional control at the protein level may have evolved to harness fully the activities of this potent oncogenic mutator in higher vertebrates^{24–26}. The B-cell signalling pathways that lead to PKA activation²⁷, and whether PKA can affect AID functionality through mechanisms beyond RPA association, remain to be determined. In T cells, NF- κ B exists in an inactive cytoplasmic complex with PKA until activation via the T-cell receptor leads to PKA-dependent NF- κ B activation and transport to the nucleus, where it binds to a co-factor to effect specific transcriptional events^{27,28}. By analogy to NF- κ B, our finding that activated B-cell nuclear extracts have the highest level of AID dsDNA deamination activity—despite cytoplasmic extracts having more AID ssDNA deamination activity—raises the possibility that AID may also be sequestered in an inactive state in the B-cell cytoplasm and become activated after appropriate upregulation of PKA.

METHODS

In vitro AID phosphorylation. A total of 60 pmol of AID(293) was incubated with 50 units of recombinant PKA (Sigma) for 1 h at 30 °C in a buffer containing 20 mM TrisHCl (pH 7.5), 80 mM NaCl, 1 mM dithiothreitol, 10 mM MgCl₂ and 1 mM ATP.

Antibodies and reagents. The sources of the antibodies used are as follows: anti-PKAc_{at}, anti-PKAr and anti-RPA₃₂ are from Santa Cruz Biotechnology, anti-GFP is from Clontech, and anti-AID is as described previously⁷. The PKA inhibitor H89 was obtained from Calbiochem and the kemptide assay kit for the PKA activity system was obtained from Promega.

Protein preparation and deamination assays. AID was purified from mouse splenic B cells or 293 human kidney cells, ectopically expressing the protein, by procedures described previously^{7,13}. Deamination assays on ssDNA or dsDNA containing RGYW motifs were described previously^{7,13}.

Recombinant retroviral production and infection of B cells. Mutant AID (T27A, S38A, Y184A) was generated by mutation of AID open reading frame (ORF) cloned in pcDNA3.0 (Invitrogen) using site-directed mutagenesis (Stratagene). The mutant AID ORF was subcloned in pMX-IRES-GFP, constructed as described²⁹. The plasmid was transfected into a virus packaging cell line (Phoenix cells, Orbigen) using the Trans-It reagent (Mirus). The virus was mixed with pre-activated B cells (activated by anti-CD40 antibodies and IL-4) for 3 days at a density of 10⁶ cells ml⁻¹ supplemented with 16 μ g ml⁻¹ of polybrene. These cells were centrifuged at 2,500 r.p.m. for 30 min at 30 °C and incubated for 48–72 h. They were then assayed for surface IgG1 production using flow cytometry. B cells were isolated from wild-type or AID^{-/-} mouse spleens through negative purifications of B cells in MACS columns with anti-CD43 antibody-coated magnetic beads (Miltenyi Biotec.). The isolated B cells were incubated in RPMI medium containing 50 μ M β -mercaptoethanol with 15% fetal calf serum (FCS), and stimulated with anti-CD40 antibodies and IL-4 for 24 h before infection.

Flow cytometric analysis. The antibodies PE-Cy5-conjugated anti-mouse CD45R (B220) (eBioscience) and PE anti-mouse IgG1 (A85-1) (BD Pharmingen) were used for staining of B cells destined for flow cytometric analysis. The analysis was performed with a BD FACScalibur.

Received 30 August; accepted 26 September 2005.

Published online 26 October 2005.

- Honjo, T., Kinoshita, K. & Muramatsu, M. Molecular mechanism of class switch recombination: Linkage with somatic hypermutation. *Annu. Rev. Immunol.* **20**, 165–196 (2002).
- Manis, J. P., Tian, M. & Alt, F. W. Mechanism and control of class-switch recombination. *Trends Immunol.* **23**, 31–39 (2002).
- Chaudhuri, J. & Alt, F. W. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature Rev. Immunol.* **4**, 541–552 (2004).
- Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
- Muramatsu, M. et al. Specific expression of activation-induced cytidine

deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476 (1999).

- Revy, P. et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565–575 (2000).
- Chaudhuri, J. et al. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726–730 (2003).
- Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl Acad. Sci. USA* **100**, 4102–4107 (2003).
- Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* **197**, 1291–1296 (2003).
- Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–103 (2002).
- Ramiro, A. R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M. C. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature Immunol.* **4**, 452–456 (2003).
- Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61–92 (1997).
- Chaudhuri, J., Khuong, C. & Alt, F. W. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* **430**, 992–998 (2004).
- Zarrin, A. A. et al. An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nature Immunol.* **5**, 1275–1281 (2004).
- Tasken, K. & Aandahl, E. M. Localized effects of cAMP mediated by distinct routes of protein kinase A. *Physiol. Rev.* **84**, 137–167 (2004).
- Nakamura, M. et al. High frequency class switching of an IgM⁺ B lymphoma clone CH12F3 to IgA⁺ cells. *Int. Immunol.* **8**, 193–201 (1996).
- Manis, J. P. et al. Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *J. Exp. Med.* **187**, 2081–2089 (1998).
- Dominguez-Sola, D. & Dalla-Favera, R. PINning down the c-Myc oncoprotein. *Nature Cell Biol.* **6**, 288–289 (2004).
- Xu, Y. X., Hirose, Y., Zhou, X. Z., Lu, K. P. & Manley, J. L. Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev.* **17**, 2765–2776 (2003).
- Yang, X. J. Multisite protein modification and intramolecular signalling. *Oncogene* **24**, 1653–1662 (2005).
- Rada, C., Di Noia, J. M. & Neuberger, M. S. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell* **16**, 163–171 (2004).
- Weill, J. C. & Reynaud, C. A. RPA tightens AID to DNA...editing. *Nature Immunol.* **5**, 876–878 (2004).
- Barreto, V. M. et al. AID from bony fish catalyzes class switch recombination. *J. Exp. Med.* **202**, 733–738 (2005).
- Kotani, A. et al. A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc. Natl Acad. Sci. USA* **102**, 4506–4511 (2005).
- Ramiro, A. R. et al. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* **118**, 431–438 (2004).
- Unniraman, S., Zhou, S. & Schatz, D. G. Identification of an AID-independent pathway for chromosomal translocations between the IgH switch region and Myc. *Nature Immunol.* **5**, 1117–1123 (2004).
- Torgersen, K. M., Vang, T., Abrahamsen, H., Yaqub, S. & Tasken, K. Molecular mechanisms for protein kinase A-mediated modulation of immune function. *Cell. Signal.* **14**, 1–9 (2002).
- Zhong, H., SuYang, H., Erdjument-Bromage, H., Tempst, P. & Ghosh, S. The transcriptional activity of NF- κ B is regulated by the I κ B-associated PKAc subunit through a cyclic AMP-independent mechanism. *Cell* **89**, 413–424 (1997).
- Fagarasan, S., Kinoshita, K., Muramatsu, M., Ikuta, K. & Honjo, T. *In situ* class switching and differentiation to IgA-producing cells in the gut lamina propria. *Nature* **413**, 639–643 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Honjo for providing AID-deficient mice and a retroviral vector for expression of AID in B cells, and R. Dalla-Favera and L. Pasqualucci for communicating unpublished findings. U.B. is the Prismatic fellow of the Irvington Institute of Immunology; S.D. is supported by a fellowship from the Swiss National Science Foundation; S.R. is supported by a National Research Service Award. This work was supported by an NIH grant to F.W.A. F.W.A. is an Investigator and G.L. an Associate of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.W.A. (alt@enders.tch.harvard.edu).

LETTERS

Spatial regulation of β -actin translation by Src-dependent phosphorylation of ZBP1

Stefan Hüttelmaier^{1,2}, Daniel Zenklusen¹, Marcell Lederer², Jason Dichtenberg³, Mike Lorenz¹, XiuHua Meng¹, Gary J. Bassell³, John Condeelis¹ & Robert H. Singer¹

Localization of β -actin messenger RNA to sites of active actin polymerization modulates cell migration during embryogenesis, differentiation and possibly carcinogenesis^{1–5}. This localization requires the oncofetal protein ZBP1 (Zipcode binding protein 1), which binds to a conserved 54-nucleotide element in the 3′-untranslated region of the β -actin mRNA known as the ‘zipcode’. ZBP1 promotes translocation of the β -actin transcript to actin-rich protrusions in primary fibroblasts and neurons^{6,7}. It is not known how the ZBP1–RNA complex achieves asymmetric protein sorting by localizing β -actin mRNA. Here we show that chicken ZBP1 modulates the translation of β -actin mRNA. ZBP1 associates with the β -actin transcript in the nucleus and prevents premature translation in the cytoplasm by blocking translation initiation. Translation only occurs when the ZBP1–RNA complex reaches its destination at the periphery of the cell. At the endpoint of mRNA transport, the protein kinase Src promotes translation by phosphorylating a key tyrosine residue in ZBP1 that is required for binding to RNA. These sequential events provide both temporal and spatial control over β -actin mRNA translation, which is important for cell migration and neurite outgrowth.

In order to characterize how the ZBP1–RNA complex achieves asymmetric protein sorting, we established an *in vitro* cell model using the NG108-15 neuroblastoma cell line (NG). As in primary embryonic fibroblast or neuronal cells, ZBP1 colocalized with β -actin mRNA at the leading edge, in growth cone filopodia and along the neurites of differentiated NG cells⁸ (Fig. 1a, b, d, e). At the cell periphery, sites of colocalization were interspersed with regions where β -actin mRNA and ZBP1 were separated, suggesting release of the mRNA transcript (Supplementary Fig. 1a–c). In the nucleus, ZBP1 was found in discrete foci, coinciding with the sites of β -actin mRNA transcription and indicating that ZBP1 associates with β -actin mRNA co-transcriptionally^{9,10} (Fig. 1c, f and Supplementary Fig. 1d–f).

A well-accepted paradigm of mRNA localization is that it precedes translation. We propose that assembly of a localized mRNA–protein complex occurs in the nucleoplasm, priming the mRNA to be translationally silenced when it reaches the cytoplasm. On the basis of this hypothesis, we tested recombinant ZBP1 for a role in regulating the translation of β -actin mRNA.

In rabbit reticulocyte lysate, ZBP1 impaired translation of the β -actin transcript in a zipcode-dependent manner, without affecting RNA stability (Fig. 1g and Supplementary Fig. 1g, h). Hence, the zipcode sequence seems to be essential for controlling both mRNA localization and translation. Similar to the RNA-binding protein hnRNP-K, ZBP1 impaired the formation of 80S ribosomal complexes on regulated transcripts in rabbit reticulocyte lysate, but the amount and positioning of the 48S complex remained unaffected¹¹

(Supplementary Fig. 2a–d). These findings confirm ZBP1 as a bona fide translational regulator that interferes with translation by blocking translation initiation.

We then attempted to validate the role of ZBP1 as a zipcode-dependent translational regulator *in vivo*. Knockdown of endogenous ZBP1 by short interfering (si)RNA enhanced translation from reporter transcripts in a zipcode-dependent manner (Fig. 1h, i and Supplementary Fig. 1i). Specificity was further confirmed by transfecting an siRNA-insensitive chicken ZBP1 (ZBP1_{insens}) into NG cells. ZBP1_{insens} restored the translational repression of transcripts

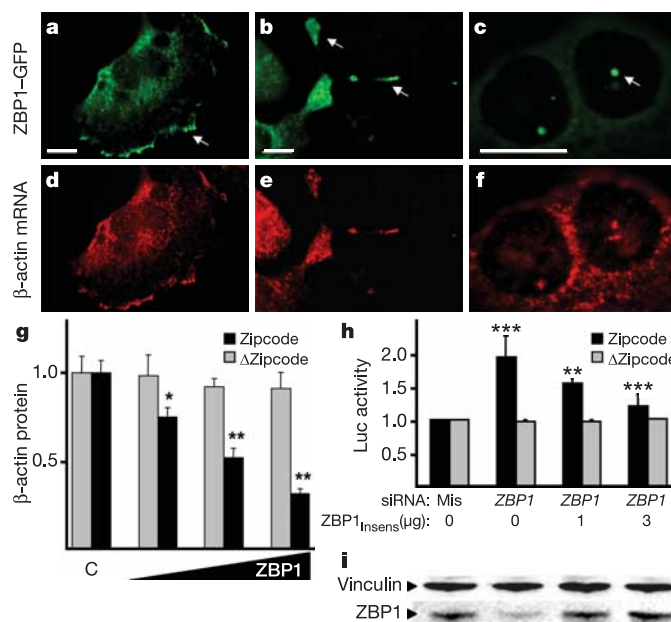


Figure 1 | ZBP1 associates with β -actin mRNA in NG cells and modulates its translation. **a–f**, Colocalization of ZBP1–GFP and β -actin mRNA in differentiated NG cells: leading edge (arrow in **a**, **d**), neurite extension (arrow in **b**, **e**) and discrete nuclear foci (arrow in **c**, **f**). Scale bars, 10 μ m. **g**, Quantification of β -actin translation in rabbit reticulocyte lysate, normalized by *Renilla* luciferase in the presence of increasing ZBP1 concentrations or buffer (C). **h**, Normalized luciferase (luc) activity from cells transfected with the indicated siRNA (Mis, missense control) and plasmids encoding increasing amounts of siRNA-insensitive ZBP1 (ZBP1_{insens}). **i**, Western blot analysis of protein extracts from **i**, using the indicated antibodies. Error bars indicate s.d. Asterisk, $P \leq 0.05$; two asterisks, $P \leq 0.005$; three asterisks, $P \leq 0.0005$.

¹Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. ²ZAMED, Department of Medicine, Martin-Luther-University of Halle, Heinrich-Damerow-Strasse 1, 06120 Halle, Germany. ³Department of Neuroscience, Rose Kennedy Center for Mental Retardation, Albert Einstein College of Medicine, 1410 Pelham Parkway, Bronx, New York 10461, USA.

containing the human β -actin mRNA zipcode (Fig. 1h, i). Hence, ZBP1 is a conserved *trans*-acting factor that controls translation of β -actin mRNA *in vivo*.

To achieve asymmetric protein sorting in neurites, a regulatory mechanism must exist to abrogate translational repression by ZBP1 once the mRNA has been localized. Sequence analysis of the ZBP1 protein reveals a possible SH3-binding motif (VP₄SS) in close proximity to a KH-domain essential for binding to the zipcode⁷. The identical motif is found in the cell-adhesion protein paxillin, where it serves as a docking site for Src kinase¹². Like paxillin, we confirmed that ZBP1 is a substrate for Src *in vivo*. Exogenous ZBP1 was tyrosine-phosphorylated by endogenous Src. Overexpression of constitutively active Src (KP, which contains a K249Q and P250G mutation) increased ZBP1 phosphorylation, whereas inactivated Src (containing a Y416F mutation) had no effect (Fig. 2a). *In vitro* and *in vivo* phosphorylation experiments identified Tyr 396 as the ZBP1 residue phosphorylated by Src (Fig. 2b, Supplementary Fig. 3 and Supplementary Fig. 4a, b). Phosphorylation of endogenous ZBP1 *in vivo* and the recombinant protein *in vitro* was verified with a specific anti-peptide antibody directed against phosphorylated Tyr 396 (Supplementary Fig. 4c, d).

Phosphorylation of tyrosine residues outside the RNA-binding domain has been suggested to decrease RNA binding of the KH-domain-containing proteins hnRNP-K and Sam68 (refs 13, 14). To address the role of Src in regulating ZBP1 function, phosphorylated ZBP1 and non-phosphorylatable ZBP1 (containing a Y396F mutation) were probed for subcellular localization and zipcode binding. No obvious differences between the wild-type and mutant protein were found (data not shown). However, zipcode binding was significantly decreased as a result of ZBP1 tyrosine phosphorylation (Supplementary Fig. 5a–c). In contrast to the wild-type protein, ZBP1-Y396F retained significantly higher binding efficiency, suggesting that phosphorylation of Tyr 396 interferes with RNA binding (Fig. 2c). ZBP1–RNA complexes isolated from cells and analysed by quantitative polymerase chain reaction with reverse transcription (RT–PCR) revealed that the association of endogenous β -actin mRNA with ZBP1 was unaffected by the Y396F mutation. However, replacement of Tyr 396 with glutamate significantly impaired the association of ZBP1 with endogenous β -actin mRNA (Fig. 2d). This indicates that tyrosine phosphorylation or the mimicking of phosphorylation by glutamate substitution interferes with the binding of β -actin mRNA both *in vitro* and *in vivo*.

The control of β -actin mRNA binding through tyrosine phosphorylation suggests that Src might also regulate translational repression through ZBP1. This was supported by our finding that increased Src activity leads to increased translation only from zipcode reporter mRNA *in vivo* (Fig. 2e). Cells that were depleted of endogenous ZBP1 by siRNA showed significantly increased translation, whereas replacement with ZBP1_{insens} or ZBP1-Y396F, which are both insensitive to siRNA, decreased translation (Fig. 2f). Notably, the mutant was slightly more efficient at reducing translation. Amino acid replacement of Tyr 396 apparently counteracts the release of β -actin mRNA, as RNA binding was unaffected by the Y396F mutation both *in vitro* and *in vivo*. Together, our data suggest that Src phosphorylation of Tyr 396 abolishes translational repression by decreasing ZBP1 binding to β -actin mRNA.

In order for translational repression to be spatially regulated, Src should associate with and phosphorylate ZBP1 at the endpoint of β -actin mRNA transport. We therefore investigated the interaction between Src and ZBP1 using immunoprecipitation experiments (Supplementary Fig. 6a). This association was impaired by introducing a negative charge C-terminal to the conserved poly-proline motif, which has been shown to interfere with SH3-mediated binding¹⁵ (Supplementary Fig. 6b–d).

We next analysed the spatial organization of the ZBP1–Src interaction by fluorescence resonance energy transfer (FRET; Fig. 3a–e). Acceptor photobleaching using ZBP1–cyan fluorescent

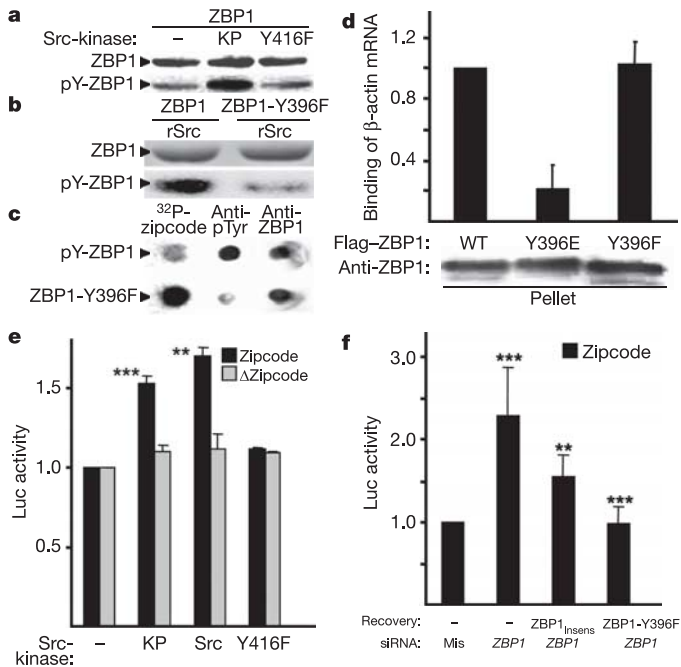


Figure 2 | Phosphorylation of ZBP1 by Src interferes with RNA binding and relieves translational repression. **a**, Western blot of Flag–ZBP1 from cells transfected with pcDNA3.1 (–), constitutively active (KP) or inactive (Y416F) Src, using the indicated antibodies. **b**, **c**, ZBP1 and ZBP1-Y396F phosphorylation by recombinant Src (rSrc), assayed by Coomassie staining (ZBP1), autoradiography (pY-ZBP1), zipcode-binding (³²P-zipcode) and immunoblotting (anti-pTyr). Anti-ZBP1 staining used as a loading control. **d**, β -actin mRNA binding to precipitated ZBP1 proteins (western blot showing precipitated ZBP1 levels), normalized to wild-type (WT) ZBP1. **e**, Luciferase activities of cells co-transfected with the indicated reporters and Src constructs, normalized relative to controls without kinase transfection. **f**, Luciferase activities relative to controls (Mis); co-transfected with the indicated siRNAs, WT, siRNA-insensitive ZBP1_{insens} or mutant ZBP1-Y396F. Error bars indicate s.d. Two asterisks, $P \leq 0.005$; three asterisks, $P \leq 0.0005$.

protein (CFP) and Src–yellow fluorescent protein (YFP) fusion proteins as donor and acceptor molecules, respectively, revealed a significant energy transfer and thus an association between Src and ZBP1 in close proximity to filopodia and in the growth cones of differentiated NG cells (Fig. 3e, arrows and arrowhead, respectively). FRET signal was not detected outside the photobleached region or when using free YFP (Supplementary Fig. 7a–f). The spatial organization of the Src–ZBP1 interaction resembled the colocalization of β -actin mRNA and ZBP1 along dendrites and within the growth cone filopodia of primary hippocampal neurons^{3,16}.

We next investigated whether disruption of translational regulation by ZBP1 had a phenotypic consequence. Previous work has suggested that β -actin mRNA localization might regulate the density of F-actin in dendritic protrusions, thereby implying that the spatial regulation of translation can modulate neuronal outgrowth by providing monomeric actin and other actin-related proteins^{8,17}. As predicted, the scoring of differentiated NG cells on the basis of cellular extension length showed that knockdown of ZBP1 significantly impaired neurite outgrowth (Fig. 3f–h and Supplementary Fig. 8a–f). Neuronal outgrowth recovered upon expression of ZBP1 that was insensitive to siRNA. In contrast, expression of non-phosphorylatable ZBP1-Y396F, which is impaired in releasing the mRNA, did not restore neuronal outgrowth. Similarly, overexpression of ZBP1-Y396F interfered with neuronal outgrowth in primary hippocampal neurons, supporting the physiological relevance of the observed phenotype (Fig. 3i–k).

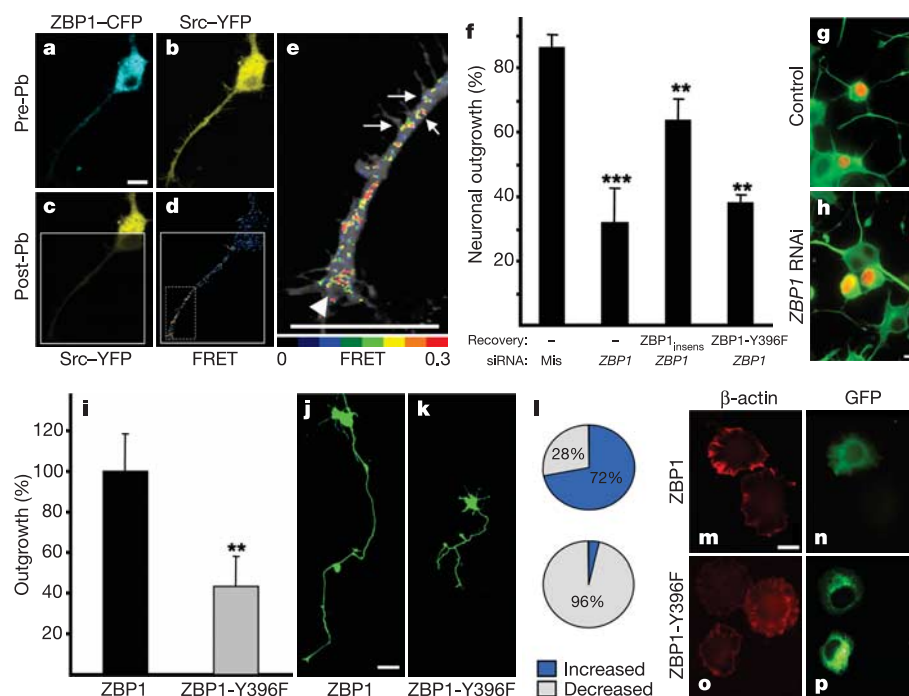


Figure 3 | Regulation of ZBP1 by Src modulates neuronal outgrowth.

a–d, FRET CFP signals before (**a**, **b**) and after YFP-photobleaching (Pb) (box in **c** and **d**). **e**, Increased FRET intensity (dashed boxed in **d**) at filopodia (arrows) and growth cone (arrowhead). **f**, Neuronal outgrowth in NG cells transfected with indicated siRNAs, siRNA-insensitive ZBP1 or ZBP1-Y396F and polypyrimidine tract binding protein (PTB)–CFP transfection marker ($n \geq 350$). **g**, **h**, PTB–CFP (red) and tubulin (green) from NG cells transfected with ZBP1 siRNA. **i–k**, Neuronal outgrowth in primary

hippocampal neurons overexpressing ZBP1-Y396F (**k**) was normalized to wild-type ZBP1 expression (**j**, $n = 40$). GFP was used as a marker for transfection. **l–p**, Per cent NG cells transfected with ZBP1–GFP (**m**, **n**) or ZBP1-Y396F–GFP (**o**, **p**), showing increased or decreased peripheral β -actin protein. β -actin (red; **m**, **o**) and GFP signal (green; **n**, **p**) signals are shown. See Supplementary Fig. 9. Scale bars, 10 μ m. Error bars indicate s.d. Asterisk, $P \leq 0.05$; two asterisks, $P \leq 0.005$; three asterisks, $P \leq 0.0005$.

We would expect that this suppression of outgrowth results directly from the inability to locally translate β -actin mRNA, as the non-phosphorylatable mutant should act as a dominant negative by failing to release translational repression at the endpoint of mRNA transport. In order to test this prediction, we assessed the levels of peripheral β -actin protein in spreading NG cells upon overexpression of ZBP1–GFP or ZBP1-Y396F–GFP (Fig. 3l–p and Supplementary Fig. 9a–h). β -Actin protein at the cell periphery was detected by indirect immunofluorescence using a β -actin-specific antibody (see Supplementary Fig. 9a–h). Enhanced expression of wild-type ZBP1 increased the actin signal at the leading edge of most spreading cells ($\sim 70\%$ of scored cells). In contrast, exogenous expression of ZBP1-Y396F reduced peripheral actin intensity in over 90% of the scored cells. These observations indicate that the cellular amount of peripheral β -actin protein is regulated through tyrosine phosphorylation of ZBP1. Therefore, spatial control of β -actin protein levels results in proper neuronal differentiation.

Our findings emphasize the temporal and spatial relationship between mRNA localization and translation. If β -actin mRNA translation were not silenced, polysome formation would initiate immediately upon nuclear egress. This would make translocation of the mRNA problematic, as polysomes are essentially immobile within the isotropic network of cytoskeletal structures¹⁸. Although the association with nuclear factors may be important for preventing β -actin translation in the cytoplasm^{19–23}, equally important is the mechanism that relieves this repression at the right time and place. Hence, ZBP1 must be removed from β -actin mRNA at the endpoint of transport. We propose that this occurs when ZBP1 comes into proximity with Src, as Src phosphorylation of ZBP1 disrupts RNA binding and activates translation. Given that Src activity is restricted to the cell periphery, local activation of β -actin translation can be

modulated solely through the spatially restricted activity of the kinase^{24,25}. Our findings indicate that a sequence of regulatory events including (1) ‘nuclear priming’, (2) translational silencing that allows cytoplasmic sorting of transcripts, and (3) spatially controlled translational de-repression of a localized transcript have evolved as a common mechanism that can have profound physiological consequences for cellular asymmetry.

METHODS

Cell culture, transfection and antibodies. NG108-15 cells and 293 cells were cultured in DMEM medium (with 10% fetal bovine serum), containing 1% HAT (hypoxanthine, aminopterin, thymidine) supplement for the culturing of NG-hybridomas. NG cells were differentiated by reducing serum levels (to 1%) and plating on a poly-L-lysine/laminin matrix. For details see Supplementary Methods.

Protein purification and *in vitro* phosphorylation. Recombinant protein was purified essentially as described⁷. Purified proteins were phosphorylated with Src (p60c-Src, Upstate) for 10–30 min at 30 °C, according to the manufacturer's instructions. After phosphorylation, proteins were immobilized on a nitrocellulose membrane before probing for RNA binding using radiolabelled zipcode, comprising the first 233 nucleotides of the human β -actin 3'–UTR.

Immunoprecipitation. Phosphoprotein or protein–RNA complexes were isolated using anti-Flag beads (Sigma) before western blotting or quantitative RT–PCR, respectively. For details see Supplementary Methods.

Molecular cloning and siRNA. The human β -actin sequence (accession number CR625492, NCBI database) or a Δ zipcode mutant (with nucleotides 1209–1441 removed) were cloned into pcDNA3.1 for *in vitro* transcription. For the luciferase reporter, the 3' region of the β -actin cDNA sequence (zipcode, nucleotides 877–1808; Δ zipcode, nucleotides 877–1208) was fused 3' of the firefly luciferase coding sequence in pcDNA3.1. ZBP1-specific siRNAs were directed against a target sequence consisting of nucleotides 393–411 of human ZBP1 (also known as IMP-1 in human, accession number AF117106, NCBI database). For siRNA-insensitive ZBP1 mutants (ZBP1_{insens}), wobble bases were

converted using site-directed mutagenesis without affecting the primary protein sequence.

FRET analysis. FRET measurements were made by acceptor photobleaching and imaging with a confocal laser scanning microscope (TCS SP2 AOBs, Leica). See Supplementary Methods for details.

Received 26 April; accepted 8 August 2005.

- Lawrence, J. B. & Singer, R. H. Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* **45**, 407–415 (1986).
- Kislauskis, E. H., Li, Z., Singer, R. H. & Taneja, K. L. Isoform-specific 3'-untranslated sequences sort α -cardiac and β -cytoplasmic actin messenger RNAs to different cytoplasmic compartments. *J. Cell Biol.* **123**, 165–172 (1993).
- Zhang, H. L. *et al.* Neurotrophin-induced transport of a β -actin mRNA complex increases β -actin levels and stimulates growth cone motility. *Neuron* **31**, 261–275 (2001).
- Shestakova, E. A., Wyckoff, J., Jones, J., Singer, R. H. & Condeelis, J. Correlation of β -actin messenger RNA localization with metastatic potential in rat adenocarcinoma cell lines. *Cancer Res.* **59**, 1202–1205 (1999).
- Shestakova, E. A., Singer, R. H. & Condeelis, J. The physiological significance of β -actin mRNA localization in determining cell polarity and directional motility. *Proc. Natl Acad. Sci. USA* **98**, 7045–7050 (2001).
- Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L. & Singer, R. H. Characterization of a β -actin mRNA zipcode-binding protein. *Mol. Cell Biol.* **17**, 2158–2165 (1997).
- Farina, K. L., Hüttelmaier, S., Musunuru, K., Darnell, R. & Singer, R. H. Two ZBP1 KH domains facilitate β -actin mRNA localization, granule formation, and cytoskeletal attachment. *J. Cell Biol.* **160**, 77–87 (2003).
- Eom, T., Antar, L. N., Singer, R. H. & Bassell, G. J. Localization of a β -actin messenger ribonucleoprotein complex with zipcode-binding protein modulates the density of dendritic filopodia and filopodial synapses. *J. Neurosci.* **23**, 10433–10444 (2003).
- Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts *in situ*. *Science* **280**, 585–590 (1998).
- Oleynikov, Y. & Singer, R. H. Real-time visualization of ZBP1 association with β -actin mRNA during transcription and localization. *Curr. Biol.* **13**, 199–207 (2003).
- Ostareck, D. H. *et al.* Lipoygenase mRNA silencing in erythroid differentiation: The 3' UTR regulatory complex controls 60S ribosomal subunit joining. *Cell* **104**, 281–290 (2001).
- Schaller, M. D. Paxillin: a focal adhesion-associated adaptor protein. *Oncogene* **20**, 6459–6472 (2001).
- Ostareck-Lederer, A. *et al.* c-Src-mediated phosphorylation of hnRNP K drives translational activation of specifically silenced mRNAs. *Mol. Cell Biol.* **22**, 4535–4543 (2002).
- Lukong, K. E. & Richard, S. Sam68, the KH domain-containing superSTAR. *Biochim. Biophys. Acta* **1653**, 73–86 (2003).
- Mayer, B. J. SH3 domains: complexity in moderation. *J. Cell Sci.* **114**, 1253–1263 (2001).
- Tiruchinapalli, D. M. *et al.* Activity-dependent trafficking and dynamic localization of zipcode binding protein 1 and β -actin mRNA in dendrites and spines of hippocampal neurons. *J. Neurosci.* **23**, 3251–3261 (2003).
- Mingle, L. A. *et al.* Localization of all seven messenger RNAs for the actin-polymerization nucleator Arp2/3 complex in the protrusions of fibroblasts. *J. Cell Sci.* **118**, 2425–2433 (2005).
- Medalia, O. *et al.* Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* **298**, 1209–1213 (2002).
- Huynh, J. R., Munro, T. P., Smith-Litiere, K., Lepesant, J. A. & Johnston, D. S. The *Drosophila* hnRNP A/B homolog, Hrp48, is specifically required for a distinct step in *osk* mRNA localization. *Dev. Cell* **6**, 625–635 (2004).
- Farina, K. L. & Singer, R. H. The nuclear connection in RNA transport and localization. *Trends Cell Biol.* **12**, 466–472 (2002).
- Yano, T., de Quinto, S. L., Matsui, Y., Shevchenko, A. & Ephrussi, A. Hrp48, a *Drosophila* hnRNP A/B homolog, binds and regulates translation of *oskar* mRNA. *Dev. Cell* **6**, 637–648 (2004).
- Hoek, K. S., Kidd, G. J., Carson, J. H. & Smith, R. hnRNP A2 selectively binds the cytoplasmic transport sequence of myelin basic protein mRNA. *Biochemistry* **37**, 7021–7029 (1998).
- Kress, T. L., Yoon, Y. J. & Mowry, K. L. Nuclear RNP complex assembly initiates cytoplasmic RNA localization. *J. Cell Biol.* **165**, 203–211 (2004).
- Bromann, P. A., Korkaya, H. & Courtneidge, S. A. The interplay between Src family kinases and receptor tyrosine kinases. *Oncogene* **23**, 7957–7968 (2004).
- Playford, M. P. & Schaller, M. D. The interplay between Src and integrins in normal and tumour biology. *Oncogene* **23**, 7928–7946 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. and D. Ostareck-Lederer for providing Src plasmids and for helpful discussions. We also thank S. Shenoy for help with microscopy and image processing. This research was supported by NIH grants to R.H.S., G.J.B. and J.C., a DOE grant to R.H.S., and by a BMBF (Bundesministerium für Bildung und Forschung) grant to S.H.

Author Contributions All of the work in the manuscript was done by S.H., with help from D.Z., X.M. and M. Lederer in the Singer laboratory, and most recently the Hüttelmaier laboratory, except for the work with primary neurons (Fig. 3i–k; J.D. and G.J.B.) and the FRET studies (Fig. 3a–e and Supplementary Fig. S7a–f; M. Lorenz and J.C.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.H.S. (rhsinger@aecon.yu.edu) or S.H. (stefan.huettelmaier@medizin.uni-halle.de).

LETTERS

Structures of ParB bound to DNA reveal mechanism of partition complex formation

Maria A. Schumacher^{1†} & Barbara E. Funnell²

The faithful inheritance of genetic information, which is essential for all organisms, requires accurate DNA partition (segregation) at cell division. In prokaryotes, partition is mediated by *par* systems, for which the P1 plasmid system of *Escherichia coli* is a prototype comprising a partition site and two proteins, ParA and ParB^{1,2}. To form the partition complex necessary for segregation, P1 ParB must recognize a complicated arrangement of A-box and B-box DNA motifs located on opposite ends of a sharply bent *parS* partition site of ~74 bp (refs 3–7). Here we describe structures of ParB bound to partition sites. ParB forms an asymmetric dimer with extended amino-terminal HTH (helix–turn–helix) domains that contact A-boxes. The two HTH domains emanate from a dimerized DNA-binding module composed of a six-stranded β -sheet coiled-coil that binds B-boxes. Strikingly, these individual DNA-binding modules rotate freely about a flexible linker, enabling them to contact several arrangements of A- and B-boxes. Most notably, each DNA-binding element binds to and thus bridges adjacent DNA duplexes. These unique structural features of ParB explain how this protein can bind complex arrays of A- and B-box elements on adjacent DNA arms of the looped partition site.

Despite the essential role of accurate DNA segregation in bacterial cell growth, the molecular mechanisms used by partition proteins in bacterial and plasmid segregation are not well understood. The *Escherichia coli* P1 plasmid partition system has served as a model of partition. A crucial step in partition is the initial formation of the so-called 'partition complex' between P1 ParB and the partition site. Partition complexes, which are thought to pair between plasmids⁸, are then bound by the ATPase ParA, which delivers each plasmid to its correct cellular location^{6,9}. ParB of *E. coli* P1 consists of a proteolytically sensitive N-terminal region (residues 1–141), which binds ParA and forms higher order oligomers, and a carboxy-terminal dimerization region (residues 142–333), which contains all of the determinants required for DNA binding¹⁰.

The *parS-small* site is the minimal partition site required for segregation¹¹ (Fig. 1a). However, partition efficiency is increased when full-length *parS* is used and the site is bent by the host auxiliary factor IHF³ (Fig. 1a). Intrinsically bent DNA can substitute for IHF, confirming that its role is simply to bring together the *parS* arms containing the A- and B-boxes^{4,12,13}. Once the initial complex is formed, additional ParB molecules load onto the DNA to form large nucleoprotein complexes^{5,6,14,15}. These findings indicate a highly complex P1 partition interaction topology in which ParB, in an unknown way, bridges the juxtaposed A- and B-box-containing arms of the looped *parS* site. To understand the DNA-binding mechanism used by ParB to form this complicated partition apparatus, we have determined crystal structures of ParB(142–333) bound to the *parS-small* partition site (Fig. 1b, c, and Methods).

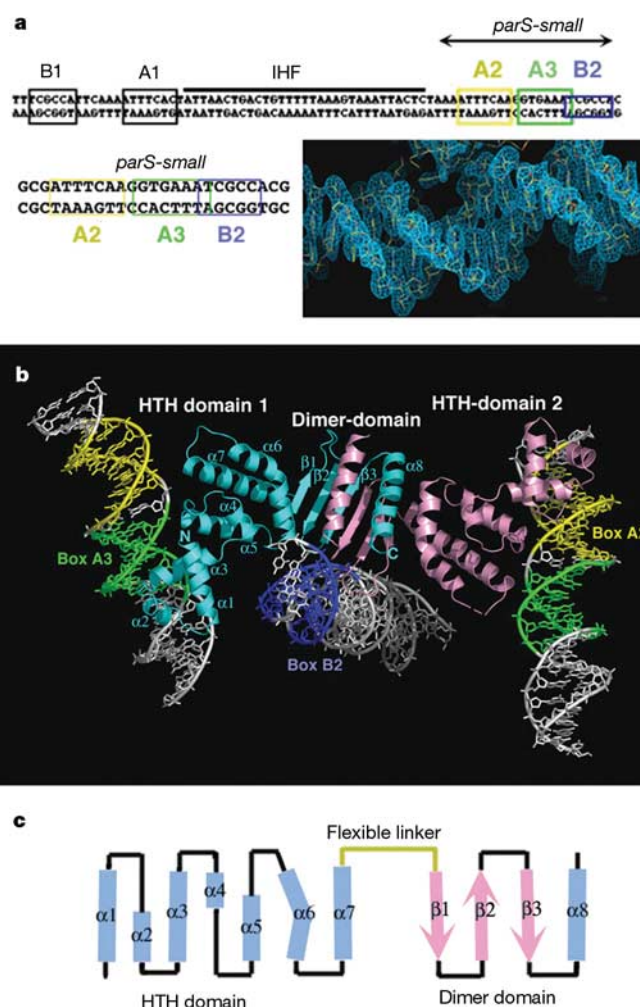


Figure 1 | Overall structure of the ParB-*parS-small* complex. **a**, Sequence of full-length P1 *parS*. A2, A3 and B2 are boxed in yellow, green and blue, respectively. Below full-length *parS* is the *parS-small* sequence with a section of the MAD map, contoured at 1.5 σ , on the right. **b**, Structure of the ParB-*parS-small* complex. The ParB(142–333) subunits are coloured cyan and pink, and secondary-structural elements are labelled for one subunit. A2, A3 and B2 are coloured as in **a**. **c**, Topology of the ParB(142–333) subunit: α 1, 151–163; α 2, 168–174; α 3, 179–189; α 4, 193–197; α 5, 207–219; α 6, 227–244; α 7, 250–270; β 1, 277–283; β 2, 290–296; β 3, 299–306; α 8, 309–325. The elements begin with 1 to denote the first helix or strand of the ParB DNA-binding domain. Note that α 6 is kinked by a central proline.

¹Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, Oregon 97239-3098, USA. ²Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [†]Present address: M. D. Anderson Cancer Center, Department of Biochemistry & Molecular Biology, Unit 1000, 1515 Holcombe Boulevard, Houston, Texas 77030, USA.

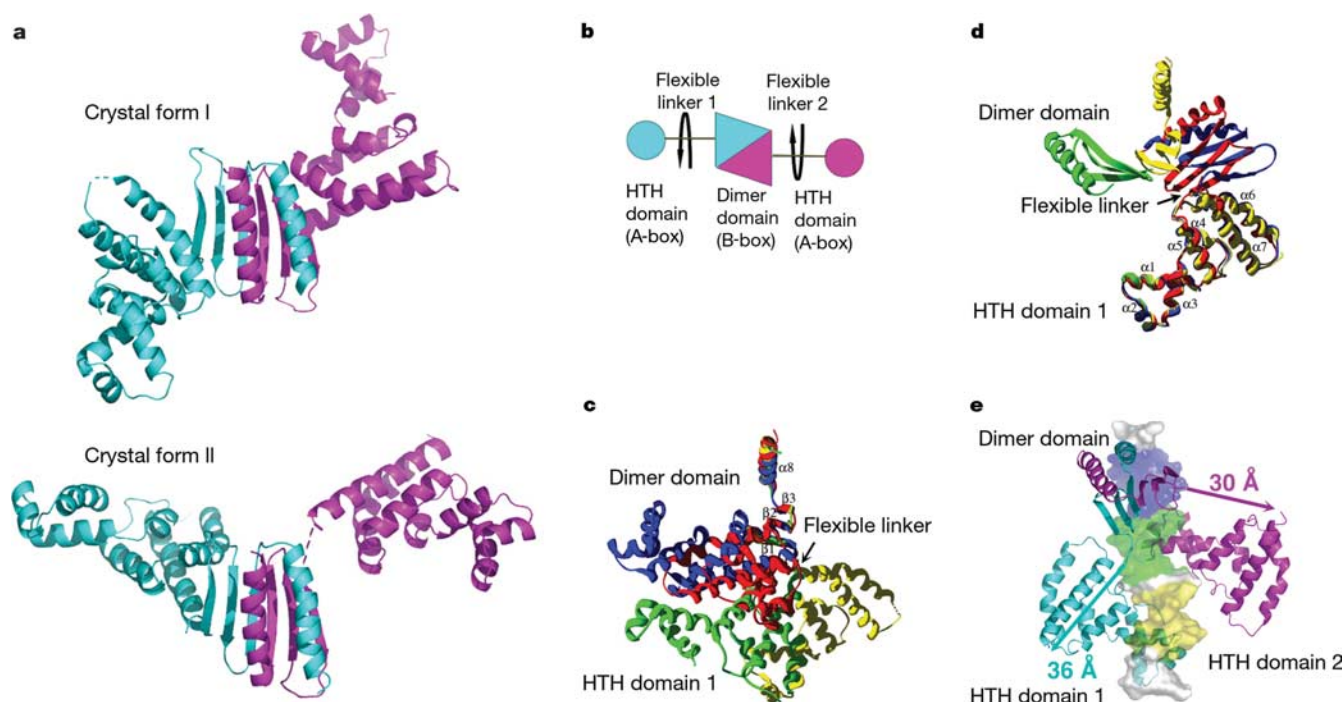


Figure 2 | ParB dimer flexibility and asymmetry. **a**, ParB dimer structure from crystal forms I and II with the dimer domains identically arranged and subunits coloured magenta and cyan. **b**, Representation of the ParB dimer (coloured as in **a**) with the HTH domains (circles) attached by the flexible linker to the dimer domain (triangles). **c**, Superimposition of the dimer

domains (indicated by underline) of the four subunits from the two crystal forms, coloured red, green, blue and yellow. **d**, Superimposition of the HTH domains, as in **c**. **e**, Model of individual DNA-binding domains bound to the same *parS-small* duplex, showing that one ParB dimer cannot bind all motifs on one *parS-small* site.

Two structures of ParB(142–333) and *parS-small* (crystal form I and II) were obtained with DNA duplexes of 25 and 22 bp containing the *parS-small* site. Each crystallographic asymmetric unit (ASU) contains one ParB(142–333) dimer, one A2 box, one A3 box and one B2 box. Crystal form I was solved by multiple-wavelength anomalous diffraction (MAD) and form II by molecular replacement (Methods). Both structures show a ParB(142–333) dimer that is simultaneously bound to three DNA duplexes. ParB(142–333) consists of two independent domains that are flexibly linked and, in the two crystal forms, the domains show marked ranges of rotations that enable them to contact DNA duplexes in distinct orientations (Figs 1b, 2 and 3). The two domains of ParB(142–333), which are

identical in the two crystal forms, are connected by a flexible linker comprising residues 271–274 and consist of an N-terminal domain, an all-helical region (the HTH domain, residues 147–270) and a dimerization domain (the ‘dimer domain’, residues 275–333)^{16,17} (Fig. 1b, c, and Supplementary Fig. 1). No stabilizing interactions are found between the two domains, which function as completely independent DNA-binding modules.

The HTH domain contains seven helices: $\alpha 2$ – $\alpha 3$ form the HTH DNA-binding unit, and the distorted helical bundle comprising $\alpha 4$ – $\alpha 7$ attaches, by the flexible linker, to the dimer domain. Each individual dimer domain consists of three antiparallel β -strands and a C-terminal α -helix (Fig. 1b, c). In the dimer, which buries 1,513 Å²

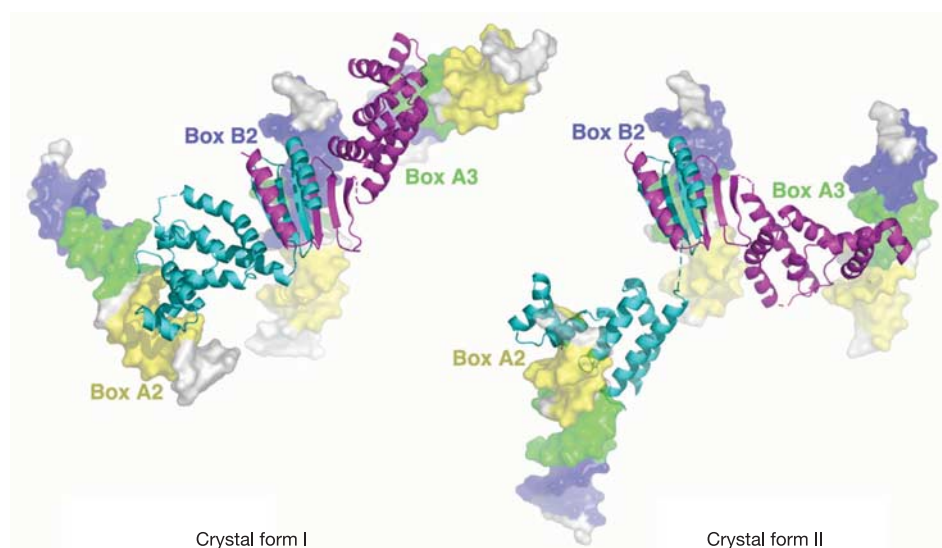


Figure 3 | P1 ParB is a DNA-bridging factor. Ribbon diagram showing the four distinct DNA bridging interactions observed in crystal forms I and II. ParB and box motifs are coloured as in Fig. 1b. The DNA is shown as a transparent surface. ParB is shown in the same orientation in forms I and II to underscore the ability of the DNA-binding modules to undergo essentially free rotation, enabling them to contact distinct orientations of box elements on different DNA duplexes.

of monomer surface, the β -strands lock together to form a continuous six-stranded antiparallel β -sheet and the C-terminal helices, which buttress the β -sheet, combine to form an antiparallel coiled-coil (Figs 1b and 2a). A marked and functionally important feature of the ParB dimer, suggested by previous biophysical studies, is its asymmetry, which arises from the flexible attachment of the domains¹³ (Figs 2a, b, and 3). In fact, the structures in the two crystal forms, which reveal domain rotations ranging from $\sim 60^\circ$ to 160° relative to one another, suggest that the flexible linker facilitates essentially free rotation of the modules (Fig. 2b–d).

Unique features of the ParB–DNA interaction highlighted in the structures explain how ParB can contact arrays of box elements located on a looped DNA site. Each box element is recognized by a separate ParB DNA-binding motif: the A-boxes are bound by the HTH domains, and the B-boxes by the dimerized dimer domains (Figs 1b and 3). Because these DNA-binding modules are flexibly attached, they can rotate freely to contact various arrangements of the A and B-boxes (Fig. 2a–d). The most striking aspect of the ParB–DNA structures, however, is the finding that each DNA-binding module in a given ParB dimer binds a box element located on a different, adjacent duplex. Thus, ParB acts as a DNA-bridging factor (Figs 1b and 3). In fact, the individual DNA-binding modules of one ParB dimer cannot simultaneously contact A- and B-boxes located on the same DNA duplex because the domain connections would be ≥ 30 Å from each other, which is too far for the four-residue linker to extend (Fig. 2e). Notably, the unique bridging function of ParB may be involved in directing plasmid pairing by enabling ParB to bridge between the two arms of one *parS* site and simultaneously to bind to a second plasmid *parS* site (Figs 1b and 3).

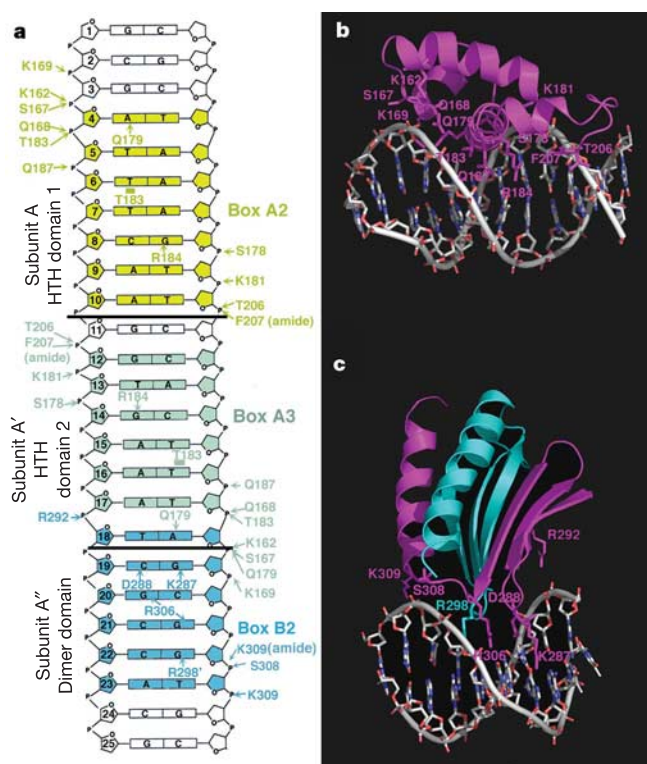


Figure 4 | ParB interactions with A- and B-box motifs. **a**, Representation of ParB contacts to the 25-bp *parS*-small site. One strand is numbered 1–25, the other is numbered 1'–25'. The 22-bp site, from crystal form II, contains the same sequence but includes only base pairs 3–24. The A2, A3 and B2 boxes are coloured as in Fig. 1b. Each box motif is bound by a domain from a separate dimer, as indicated by subunits A, A' and A''. Hydrophobic contacts are indicated by rectangles and other contacts by arrows. **b**, Contacts between the HTH domain and the A-box. **c**, Contacts between the dimer domain and the B-box.

The two HTH domains of each ParB(142–333) dimer dock in the same way and make identical contacts to A2 and A3 boxes located on different duplexes. Similar to most canonical HTH motifs, the ParB HTH motif uses residues from its recognition helix, $\alpha 3$, to make specific contacts with A-box major groove bases (Fig. 4a, b). A striking feature of the structure is the use of the dimerized dimer domain to mediate sequence-specific DNA binding. In fact, several studies have shown that ParB dimerization is important for DNA binding^{10,18}. The structure reveals that this domain must dimerize to form a functional DNA-binding module. Moreover, the dimerized module, which contacts the B-box, represents a previously unknown type of DNA-binding motif (Fig. 4a, c).

This module is buttressed onto the DNA by hydrogen bonds and helix dipole interactions from the N terminus of one helix of the coiled-coil. Docked in this orientation, residues from one end of the β -sheet reach deep into the major groove to make several base-specific contacts. Arg 298 from one subunit contacts the base G22'. The remaining base contacts are provided from the other subunit (Fig. 4a, c). Specifically, Arg 306 contacts the O6 atoms of bases G20 and G21' by its NH_1 and NH_2 groups, and both bases of base pair 19 are specified by Asp 288 and Lys 287: Asp 288 contacts the N4 atom of nucleotide C19, whereas Lys 287 contacts the O6 atom of base G19'. The importance of the contacts between ParB and the B-box is supported by experiments showing that the region corresponding to residues 281–290 of P1 ParB, known as the DRS ('discriminator recognition sequence'), provides crucial partition site recognition in this subclass of *par* systems^{19–21}. Consistent with the structure, only *Salmonella typhimurium* pSLT and *Salmonella flexneri* pWR501 *par* systems contain ParB proteins with the DRS residues Lys 287 and Asp 288 and recognize B-boxes with the same 5' sequences (Supplementary Fig. 2).

ParB proteins from different partition classes show little sequence homology and bind divergent centromere-like sites. However, a common feature of some ParB proteins is their organization with a central HTH domain and a C-terminal dimerization domain^{1,2,22,23} (Supplementary Information). The structures of ParB(142–333) and *parS*-small represent the first structures of a ParB protein that recognizes a complicated *par* site that includes two DNA-binding elements. For the P1 system, numerous studies have shown that, in the presence of IHF, one ParB dimer binds across *parS* to form the initial P1 partition complex, called I+B1 (refs 1, 2, 4, 5, 13). How ParB can bind between the looped *parS* arms has, however, remained unclear. The ParB–DNA structures reveal two atypical DNA-binding features of ParB, its bridging capability and its flexibly attached DNA-binding modules, which explain how this process occurs.

The structures predict that only one A-box and one B-box on each arm of *parS* are necessary to form the I+B1 bridged complex. On the basis of structural considerations, it seems unlikely that A-boxes on different arms can be simultaneously bound by the same dimer that binds one B-box (Figs 1b and 3). However, both B-boxes may be bound by one ParB dimer across the bent full-length *parS*⁴. Only one interaction between the dimer domain and the B-box is observed in our structures owing to the 1:1 stoichiometry of ParB dimer to B-box used for crystallization. But modelling shows that both faces of the dimer domain could interact with B-boxes from two different DNA duplexes arranged as direct repeats, such as in full-length *parS* (Supplementary Figs 1a and 3). Thus, the data suggest that I+B1 consists of a mixture of one ParB dimer bound across the *parS* bend to different combinations of A1, A2, A3, B1 and B2 boxes. The formation of the I+B1 complexes would represent the first step in the assembly of the P1 partition complexes whereby additional ParB dimers would load onto the DNA to form the large nucleoprotein complexes that are competent for partition.

METHODS

Protein preparation, crystallization and data collection. ParB(142–333) with a hexa-histidine tag was expressed and used for structural studies. To prepare

selenomethionine-substituted ParB(142–333), we expressed the protein using the methionine inhibitory pathway and purified it as described for wild type (Supplementary Information). More than 70 deoxyoligonucleotides containing the *parS-small* site were used for crystallization trials. Of the 41 different crystal forms obtained, only one diffracted to beyond 4.0 Å resolution. These crystals, which were characterized as orthorhombic ($P2_12_12$, $a = 60.5$ Å, $b = 184.3$ Å, $c = 71.2$ Å), were grown by the hanging-drop vapour-diffusion method by mixing 2 µl of a solution containing 1 mM ParB(142–333) dimer and 1 mM 22-bp *parS-small* duplex DNA with 2 µl of a reservoir solution containing 15% 2-methyl-2,4-pentanediol (MPD), 0.2 M KCl, 0.1 M HEPES (pH 7.5) and 10 mM MgCl₂ (Fig. 1a). However, the resolution could not be improved beyond 3.90 Å and the structure could not be solved. Therefore, an exhaustive search of conditions was undertaken to try and improve the diffraction quality of the remaining 40 crystals.

Crystal form I was grown by mixing a 1:1 solution of ParB(142–333) and the 25-bp *parS-small* duplex (Fig. 1a) with an equal volume of a reservoir solution containing 1.4 M sodium citrate and 0.1 M imidazole (pH 7.5). These crystals were hexagonal ($P6_322$, $a = b = 154.6$ Å, $c = 143.2$ Å) and diffracted to only 10.0 Å resolution at synchrotron sources. However, dehydration of these crystals (Supplementary Information) improved their diffraction limit to beyond 3.00 Å resolution. MAD and native X-ray intensity data were collected from dehydrated selenomethionine-substituted crystals at 100 K at Advanced Light Source (ALS) beamline 8.2.1. We collected a data set at 3.90 Å resolution for the $P2_12_12$ crystal form (crystal form II) at ALS beamline 5.02. All data were processed with MOSFLM (Supplementary Table 1).

Structure determination and refinement. The structure of the complex containing ParB(142–333) and the 25-bp *parS-small* duplex was solved by MAD (Supplementary Table 1). The program Solve located all of the selenium sites, and density modification with CNS produced an excellent electron density map^{24,25} (Fig. 1a). The positions of the selenomethionine residues combined with the positions of iodine atoms, determined from X-ray intensity data collected on dehydrated crystals in which several thymine residues were replaced by 5-iodouracil, allowed unambiguous tracing of the model^{25,26}. Model building was carried out in O, resulting in a final R_{free} of 29.6% (refs 25, 26). The refined structure shows excellent stereochemistry and contains one ParB(142–333) dimer (residues 146–326 of one subunit and residues 150–244, 246–271 and 275–328 of the other subunit), one complete 25-bp *parS-small* site duplex (the three box elements in ASU combine to form the single *parS-small* duplex) and eight solvent molecules²⁷. The open nature of the lattice enforced by ParB–DNA bridging interactions, as well as the flexibility of the ParB DNA-binding modules, probably explains the poor diffraction inherent in crystals of ParB(142–333) and DNA.

The ParB DNA-binding domains are flexibly attached; thus, to determine the structure of the $P2_12_12$ crystal form, complexes of the HTH domain and A-box DNA and the dimerized dimer domain and B-box DNA were used separately as search models for molecular replacement using EPMR²⁸. The $P2_12_12$ crystal form contains the ParB(142–333) dimer and the 22-bp *parS-small* duplex in the ASU, and EPMR could correctly position the two HTH domains bound to A-boxes. The dimer domain bound to the B-box structure solution was subsequently obtained. When all solutions were combined in the final model, the DNA formed the continuous 22-bp *parS-small* site. Owing to the low resolution of the $P2_12_12$ data, the model was subjected to minimal refinement in CNS, resulting in an R_{free} of 37.0% at 3.90 Å resolution. The final model contains one ParB dimer (residues 150–244, 246–271 and 276–329 of one subunit, and 150–272 and 278–327 of the other subunit) and one complete 22-bp *parS-small* site. Figures 1b, 2a, 2e, 3 and 4b, c were generated with PyMOL²⁹ (<http://www.pymol.org>). Figure 2c, d was made with Swiss-PdbViewer³⁰ and rendered with POVray v3.1 (<http://www.povray.org>).

Received 7 June; accepted 17 August 2005.

1. Funnell, B. E. & Slavcev, R. A. in *Plasmid Biology* (eds Funnell, B. E. & Phillips, G. J.) 81–103 (ASM, Washington DC, 2004).
2. Surtees, J. A. & Funnell, B. E. Plasmid and chromosome traffic control: how ParA and ParB drive partition. *Curr. Top. Dev. Biol.* **56**, 145–180 (2003).
3. Funnell, B. E. Participation of *Escherichia coli* integration host factor in the P1 plasmid partition system. *Proc. Natl Acad. Sci. USA* **85**, 6657–6661 (1988).
4. Funnell, B. E. & Gagnier, L. The P1 plasmid partition complex at *parS*: II. Analysis of ParB protein binding activity and specificity. *J. Biol. Chem.* **268**, 3616–3624 (1993).
5. Bouet, J. Y., Surtees, J. A. & Funnell, B. E. Stoichiometry of P1 plasmid partition complexes. *J. Biol. Chem.* **275**, 8213–8219 (2000).
6. Erdmann, N., Petroff, T. & Funnell, B. E. Intracellular localization of P1 ParB

protein depends on ParA and *parS*. *Proc. Natl Acad. Sci. USA* **96**, 14905–14910 (1999).

7. Davis, M. A., Martin, K. A. & Austin, S. J. Specificity switching of the P1 plasmid centromere-like site. *EMBO J.* **9**, 991–998 (1990).
8. Edgar, R., Chatteraj, D. K. & Yarmolinsky, M. Pairing of P1 plasmid partition sites by ParB. *Mol. Microbiol.* **42**, 1363–1370 (2001).
9. Fung, E., Bouet, J. Y. & Funnell, B. E. Probing the ATP-binding site of P1 ParA: partition and repression have different requirements for ATP binding and hydrolysis. *EMBO J.* **20**, 4901–4911 (2001).
10. Surtees, J. A. & Funnell, B. E. The DNA binding domains of P1 ParB and the architecture of the P1 plasmid partition complex. *J. Biol. Chem.* **276**, 12385–12394 (2001).
11. Martin, K. A., Davis, M. A. & Austin, S. J. Fine structure analysis of the P1 plasmid partition site. *J. Bacteriol.* **173**, 3630–3634 (1991).
12. Hayes, F. & Austin, S. Topological scanning of the P1 plasmid partition site. *J. Mol. Biol.* **243**, 190–198 (1994).
13. Funnell, B. E. The P1 plasmid partition complex at *parS*. The influence of *Escherichia coli* integration host factor and of substrate topology. *J. Biol. Chem.* **266**, 14328–14337 (1991).
14. Rodionov, O., Lobocka, M. & Yarmolinsky, M. Silencing of genes flanking the P1 plasmid centromere. *Science* **283**, 546–549 (1999).
15. Rodionov, O. & Yarmolinsky, M. Plasmid partitioning and the spreading of P1 partition protein ParB. *Mol. Microbiol.* **283**, 1215–1223 (2004).
16. Holm, L. & Sander, C. Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.* **27**, 244–247 (1999).
17. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
18. Lobocka, M. & Yarmolinsky, M. P1 plasmid partition: a mutational analysis of ParB. *J. Mol. Biol.* **52**, 366–382 (1996).
19. Radnedge, L., Davis, M. A. & Austin, S. J. P1 and P7 plasmid partition: ParB protein bound to its partition site makes a separate discriminator contact with the DNA that determines species specificity. *EMBO J.* **15**, 1155–1162 (1996).
20. Radnedge, L., Youngren, B., Davis, M. & Austin, S. Probing the structure of complex macromolecular interactions by homolog specificity scanning: the P1 and P7 plasmid partitions systems. *EMBO J.* **17**, 6076–6085 (1998).
21. Hayes, F. & Austin, S. J. Specificity determinants of the P1 and P7 plasmid centromere analogs. *Proc. Natl Acad. Sci. USA* **90**, 9228–9232 (1993).
22. Khare, D., Ziegler, G., Lanka, E. & Heinemann, U. Sequence specific DNA binding determined by contacts outside the helix–turn–helix motif of the ParB homolog KorB. *Nature Struct. Mol. Biol.* **11**, 656–663 (2004).
23. Leonard, T. A., Butler, P. J. & Lowe, J. Structural analysis of the chromosome segregation protein Spo0J from *Thermus thermophilus*. *Mol. Microbiol.* **53**, 419–432 (2004).
24. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure determination. *Acta Crystallogr. D* **55**, 849–861 (1999).
25. Brünger, A. T. et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
26. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
28. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D* **55**, 484–491 (1999).
29. Delano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, California, 2002).
30. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Advanced Light Source and their support staff, with special thanks to C. Ralston. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the US Department of Energy at the Lawrence Berkeley National Laboratory. This work was supported by a Burroughs Wellcome Career Development Award (to M.A.S.) and a grant from the Canadian Institutes of Health Research (to B.E.F.).

Author Information The X-ray crystallographic coordinates and structure factor files have been deposited with the Protein Data Bank under accession code 1ZX4. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.S. (schumacm@ohsu.edu).

LETTERS

An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA

T. Martin Schmeing¹, Kevin S. Huang¹, Scott A. Strobel^{1,2} & Thomas A. Steitz^{1,2,3}

The large ribosomal subunit catalyses the reaction between the α -amino group of the aminoacyl-tRNA bound to the A site and the ester carbon of the peptidyl-tRNA bound to the P site¹, while preventing the nucleophilic attack of water on the ester, which would lead to unprogrammed deacylation of the peptidyl-tRNA. Here we describe three new structures of the large ribosomal subunit of *Haloarcula marismortui* (Hma) complexed with peptidyl transferase substrate analogues that reveal an induced-fit mechanism in which substrates and active-site residues reposition to allow the peptidyl transferase reaction. Proper binding of an aminoacyl-tRNA analogue to the A site induces specific movements of 23S rRNA nucleotides 2618–2620 (*Escherichia coli* numbering 2583–2585) and 2541(2506), thereby reorienting the ester group of the peptidyl-tRNA and making it accessible for attack. In the absence of the appropriate A-site substrate, the peptidyl transferase centre positions the ester link of the peptidyl-tRNA in a conformation that precludes the catalysed nucleophilic attack by water. Protein release factors² may also function, in part, by inducing an active-site rearrangement similar to that produced by the A-site aminoacyl-tRNA, allowing the carbonyl group and water to be positioned for hydrolysis.

During protein synthesis, the ribosome polymerizes amino acids by repeatedly positioning the α -amino group of aminoacyl-tRNAs to react with the ester carbon of peptidyl-tRNAs¹. The ester link of the peptidyl-tRNA must be protected from hydrolysis when the A site is unoccupied, to prevent the production of truncated protein fragments. The peptidyl transferase centre selectively activates one nucleophile, the α -amino group of the aminoacyl-tRNA, while preventing the activated attack by another, a water molecule. The questions that arise are analogous to those posed by Koshland nearly 50 years ago in explaining why hexokinase does not catalyse the hydrolysis of ATP in the absence of glucose³: How does the ribosome protect the peptidyl-tRNA from inopportune catalysed hydrolysis during protein synthesis when the A site is empty? Why does it not activate the nucleophilic attack of water, just as it does for the α -amino group? For hexokinase Koshland proposed³, and Bennett and Steitz established⁴, that the binding of glucose induces an enzyme conformational change that is required for the reaction, and water does not produce this change.

Early biochemical experiments found that binding of an unacylated tRNA mimic in the A site promotes the hydrolysis of a peptidyl-tRNA bound to the P site by two orders of magnitude⁵. The trinucleotide CCA, which corresponds to the acceptor end of the tRNA, promotes this hydrolysis equally well, whereas CA and A have no effect on the hydrolysis rate. These data suggest an important alteration of the active site produced by binding of tRNA to the A site, for which C74 is required. More recently, Ehrenberg and colleagues⁶ have found that whereas fMet-tRNA^{fMet} bound to the 70S P site is stable for nearly 6 h, the addition of unacylated tRNA^{Phe} to the A site

stimulates deacylation as much as 300-fold, almost to the level produced by a protein release factor. Hydrolysis at this rate could lead to one premature termination in about every five proteins produced in the cell, because the A site is mostly unoccupied during elongation as a result of rapid reaction and progression to the hybrid state⁷ after rate-limiting tRNA accommodation⁸.

To understand these observations, we have determined two structures of the large ribosomal subunit complexed simultaneously with both an A-site and a P-site substrate (Fig. 1) (2.58 and 2.75 Å resolution; Supplementary Table S1). One of the A-site substrates, cytidine-cytidine-hydroxypuromycin (CChPmn), includes C74; the other, ChPmn, does not. The α -amino group is replaced by a hydroxyl group, which decreases the rate of reaction 500-fold⁹, allowing unreacted substrates to be revealed in electron density maps. The P-site substrate in both cases is cytidine-cytidine-adenosine-phenylalanine-caproic acid-biotin (CCApbc)¹⁰. We have also determined the structure (2.60 Å resolution) of the 50S subunit complexed to a novel transition-state analogue (TSA) that mimics that tetrahedral intermediate of the reaction (Fig. 1). Systematic movements of the rRNA are observed, including the rearrangement of some nucleotides that are known to be conformationally flexible^{10–12}.

The mechanism by which the ribosome protects the peptidyl-tRNA from hydrolysis can be understood by inspecting the conformation of the peptidyl-tRNA analogue bound to the large ribosomal subunit. The P-site ester is inaccessible to nucleophilic attack by water or the α -amino group in the structures of three complexes that include a P-site-bound peptidyl-tRNA fragment without a full A-site component (Supplementary Fig. S2). The structures of the 50S subunit complexed with CCApbc and the antibiotic sparsomycin¹³ (complex 1) CCApbc bound to either the P or A sites¹³ (complex 2) and CCApbc bound in the P site and ChPmn bound in the A site (complex 3) provide an excellent approximation to the structure of the ribosome bound to a P-site substrate alone. In complex 1, CCApbc is bound homogeneously to the P site. Sparsomycin sterically overlaps the A site, but it does not make the same interactions as an A-site substrate does. Complex 2 includes a significant contribution from molecules with CCApbc bound to the P site alone, but the peptidyl moiety is overlapped with density from molecules having exclusive binding of CCApbc to the A site. Simultaneous binding of the substrate to both sites is excluded by overlap between the peptidyl portions of the molecule within the exit tunnel. In complex 3, the binding of ChPmn drives CCApbc to the P site, but the lack of C74 leaves the rRNA in an uninduced state. The uninduced conformation of the rRNA in complexes 1 and 3 are fully compatible with the observed electron density for complex 2, although the mixture of A-site and P-site binding makes the electron density in this structure difficult to interpret in the absence of the other structures. Together, these structures provide clear evidence

¹Department of Molecular Biophysics and Biochemistry, ²Department of Chemistry, Yale University and ³Howard Hughes Medical Institute, New Haven, Connecticut 06520, USA.

that the complex with substrate bound only to the P site is in the uninduced state.

These complexes all show that the ester group of CCAp**cb** is both positioned and sequestered by the 23S rRNA. U2620(2585) lies on one side of the ester and A2486(2451) and C2104(2063) are positioned on the other (Fig. 2). These interactions require the carbonyl oxygen of the ester group to point towards the A site. Small-molecule studies¹⁴ show that the optimal angle for nucleophilic attack is about 105° from the plane of the ester group. Such an attack is not possible

in this ester group orientation because the required approach is blocked from both sides (Fig. 2). Furthermore, in serine proteases, a system that resembles this case more closely, the attack angle has been shown¹⁵ to be about 90°, which would require an approach trajectory that is even more unfavourable. This conformation of rRNA around the peptidyl-tRNA could therefore protect the nascent peptide from premature hydrolysis by occluding water from appropriate attack positions and is consistent with the observation (E. Youngman and R. Green, personal communication) that binding of peptidyl-tRNA

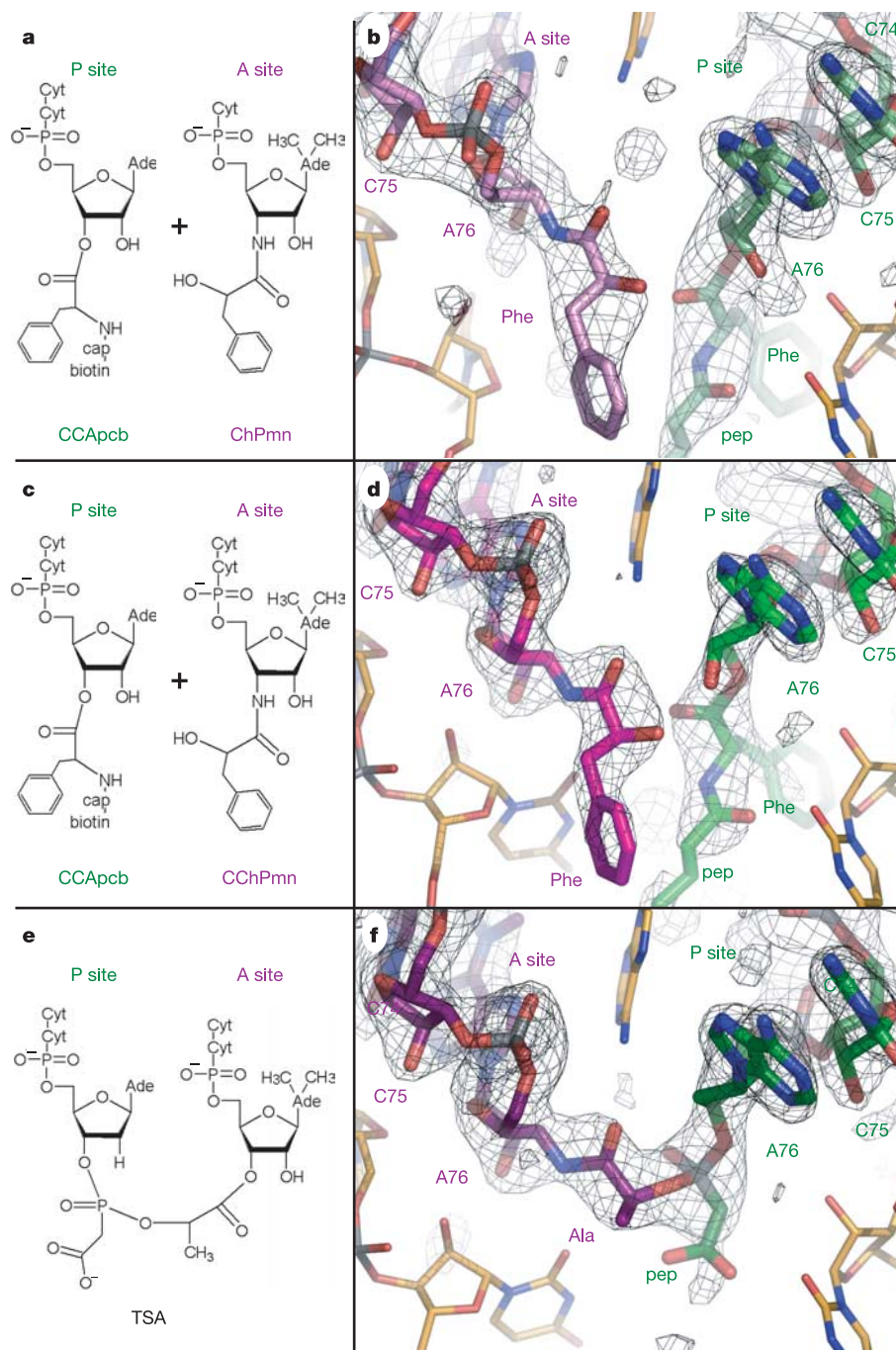


Figure 1 | Chemical structures and electron density maps. **a, c, e,** The chemical structures of CCAp**cb** and ChPmn (**a**), CCAp**cb** and CChPmn (**c**) and the TSA (**e**). **b, d, f,** Minimally biased $F_o - F_c$ electron density maps from experiments with the respective substrates: ChPmn (pink), CCAp**cb** (light green), and 23S rRNA (orange) displayed with a map calculated at 2.7 Å resolution, contoured at 2.5 σ (**b**); CChPmn (purple), CCAp**cb** (green),

and 23S rRNA (orange) displayed with map calculated at 2.4 Å resolution, contoured at 3.0 σ (**d**); the TSA (A-site portion in dark purple and the P-site portion in dark green) and 23S rRNA (orange) shown with a map calculated at 2.5 Å resolution, contoured at 3.0 σ (**f**). The same view of the peptidyl transferase centre is shown in **b, d** and **f**.

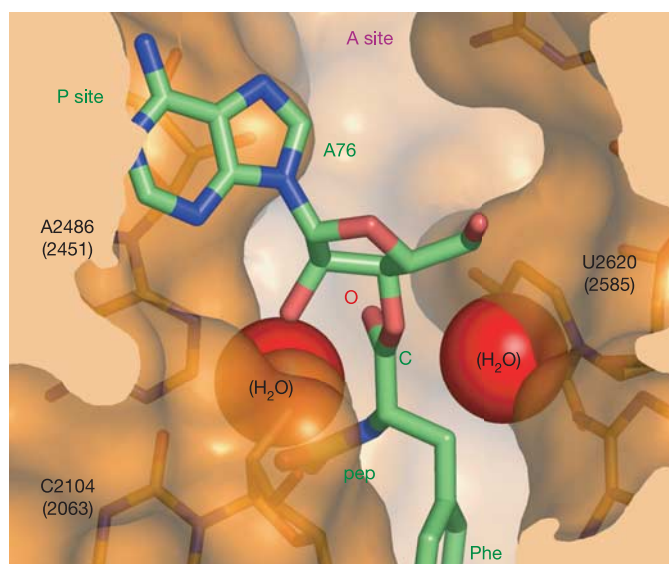


Figure 2 | Steric exclusion of water results in protection of peptidyl-tRNA from deacylation in the uninduced state. When the peptidyl transferase centre is not in the induced state, as occurs when ChPmn and CCApbc (green) are bound, the ribosome (orange surface) occludes water from positions that could attack the ester group. Theoretical water molecules (red spheres), are shown aligned for attack at 105° to the plane of the ester group, 2.8 Å away from the ester carbon. Steric clashes with A2486(2451) and C2104(2063) block the position on one side, whereas the uninduced conformation of U2620(2585) would block the other side.

to the ribosome decreases the rate of peptidyl-tRNA hydrolysis by more than 20-fold.

In the complex of the 50S subunit with ChPmn and CCApbc, not only is the P-site substrate protected from attack by a water molecule, the substrates are in a conformation that is unfavourable for peptidyl transferase reaction. The carbonyl ester oxygen of CCApbc is pointing towards the α -amino group of ChPmn, and the atoms that are to react are about 4 Å apart. The conformation of the rRNA in this

complex is similar to those with an empty A site¹⁶, which include a G–U wobble pair formed by G2618(2583) and U2541(2506). Nevertheless, ChPmn makes the interactions typical of A-site binding¹¹, as C75 of ChPmn base pairs with G2588(2553), and A76 makes an A-minor interaction¹⁷ with G2618(2583) (Fig. 3a).

Binding of an A-site substrate containing C74 induces a conformational change in the ribosome that exposes the ester group of the peptidyl-tRNA for attack by the aminoacyl-tRNA (Fig. 3, and Supplementary movies 1 and 2). The inclusion of C74 in CChPmn produces a continuous stacking interaction with U2590(2555) that positions the substrate closer to the centre of the active site (Fig. 3a). C75 of the A-site substrate still base pairs with G2588(2553), but G2618(2583) shifts to maintain the A-minor interaction with A76 (Fig. 3a). This shift of G2618(2583) causes its G–U wobble pair with U2541(2506) to break, and U2541(2506) swings through 90° (Fig. 3b). Furthermore, the movement of G2618(2583) is propagated along the rRNA chain, and the two adjacent bases, methylU2619(2584) and U2620(2585), shift in the same direction. U2620(2585), which had been near the P site with its base in van der Waals contact with the ester group of the peptidyl-tRNA mimic, swings over to approach the 2'-hydroxyl of A76 in the A site and adopts a conformation previously prevented by methylU2619(2584). This displacement of U2620(2585) creates space adjacent to the ester of the P-site substrate, allowing the peptidyl moiety and the sugar of A76 to rotate, exposing the ester to attack from the A site. The stacking of the CCA bases also lowers the α -amino group in the active site by almost 1 Å, placing it in a position that is more favourable for nucleophilic attack on the more appropriately oriented carbonyl carbon of peptidyl-CCA (Fig. 4). In this pre-reaction ground state, the hydroxyl group mimics of the α -amino forms hydrogen bonds with both N3 of A2486(2451) and the 2'-hydroxyl of A76 of the P-site substrate, a group that is critical for reaction¹⁸ and may enhance its nucleophilicity. The conformation of the rRNA in the TSA complex is nearly identical to that of the complex with CChPmn and CCApbc (Fig. 3, and Supplementary Fig. S3), showing that the RNA rearrangements induced by the binding of CCPmn are indeed likely to be relevant to the peptidyl transferase reaction pathway.

The ribosome therefore uses a substrate-induced fit mechanism to facilitate peptide bond formation and exclude hydrolysis, similar to

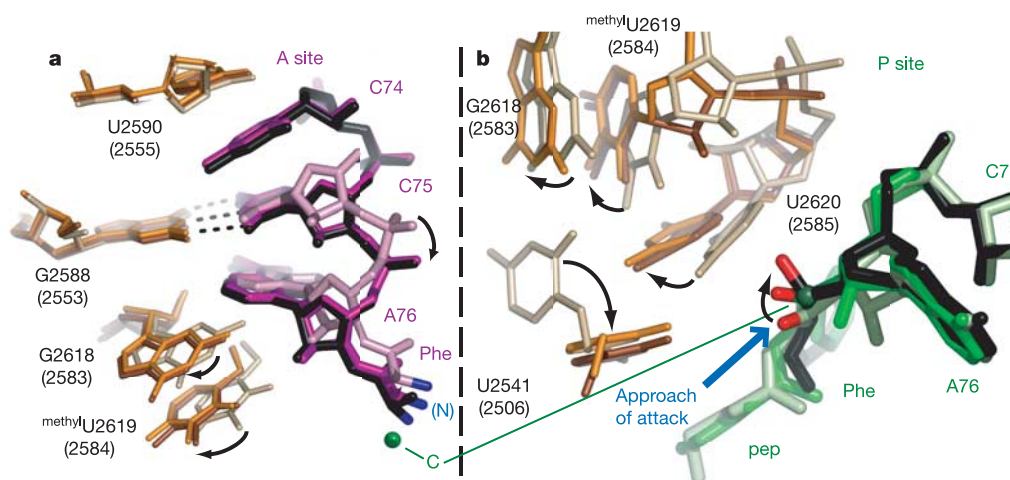


Figure 3 | Movements of rRNA and peptidyl-tRNA induced by proper binding of the A-site substrate. a, An A-site view of the three complexes. Without C74, ChPmn (pink, with rRNA in wheat colour) is positioned higher in the A site. In substrates with C74, both CChPmn (purple, with orange rRNA) and the TSA (black, with brown rRNA), C74 stacks with U2590(2555), shifting the substrates down and the α -amino group closer to the ester carbon of the P-site substrate (green). G2618(2583) shifts to maintain the A-minor interaction, causing methylU2619(2584) to move too.

b, P-site view of the same three complexes. The movement induced in G2618(2583) by A-site substrate binding breaks its G–U wobble pair with U2541(2506), which swings through 90°. methylU2619(2584) and U2620(2585) shift, allowing the ester group to move from its position in the CCApbc (light green) and ChPmn complex, to that when CCApbc (medium green) is bound with CChPmn, and finally to that when it has been attacked by the A-site substrate, as shown with the TSA (black). This induced-fit mechanism is illustrated in Supplementary movies 1 and 2.

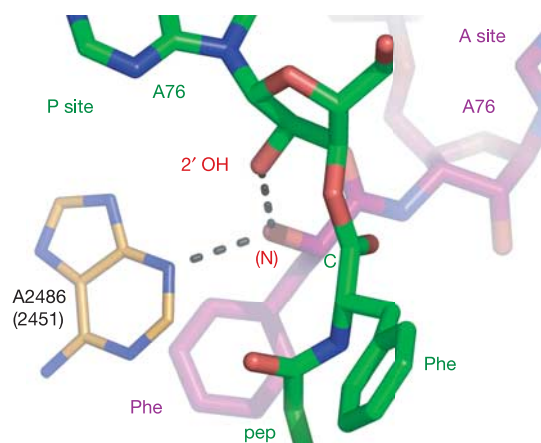


Figure 4 | Pre-attack conformation of the substrates. The hydroxyl group representing the α -amino group of the A-site substrate, CChPmn (purple) is in position to attack the ester group of the P-site substrate CCApCb (green). It is within hydrogen-bonding distance of N3 of A2486(2451) and the 2' hydroxyl group of the P-site substrate. In this ground state, the reactive groups are 3.7 Å apart.

that proposed for several enzymes, including hexokinase⁴. The reactive faces of the P-site ester are protected from nucleophilic attack by water except when a complete CCA is bound in the A site. The ester becomes exposed and properly positioned only when aminoacyl-tRNA binds the A loop, an event that is coordinated with positioning of the α -amino group for attack. If a CCA or tRNA bound to the A site is not aminoacylated, the resulting conformational change exposes the peptidyl-tRNA to nucleophilic attack by water, which accounts for the observations that either CCA⁵ or full tRNA⁶ in the A site promotes the hydrolysis of peptidyl-tRNA. Indeed, the structure of the large ribosomal subunit with CCA bound simultaneously to the A, P and E sites¹⁹ shows similar conformational changes.

Although C74 seems to be important in the proper positioning of substrate and the resultant conformational changes, the catalytic rate of peptidyl transfer is similar for cytidine-puromycin (CPmn) and CCPmn²⁰. It therefore seems that the conformational change is not rate-limiting for the fragment reaction. In the electron density map for the complex of ChPmn and CCApCb (Fig. 1), the features of the ChPmn are not as clear as expected at this resolution, indicating that there might be more than one mode of binding. Presumably both the induced and uninduced conformations are present in the crystal, but the uninduced state predominates. The apparent contradiction between the ability of CPmn to react and the inability of CA to stimulate deacylation can be attributed to an important function of the interactions made by the amino acid moiety of ChPmn in shifting the conformational equilibrium towards the induced state. CPmn can access the induced conformation to react, but the deacylated CA cannot access the induced state enough to increase the rate of hydrolysis⁵.

When a translating ribosome encounters a stop codon in mRNA, a protein release factor binds and promotes the deacylation of the peptidyl-tRNA². A universally conserved GGQ motif in class one release factors (RFs) interacts with the peptidyl transferase centre^{6,21–23}. In the context of this induced-fit model, the RF presumably promotes the conformational rearrangement of at least a subset of those rRNA nucleotides that are responsible for activation, in particular U2620(2585). Such rearrangement would allow the ester group in the P site to move out of the protected conformation and undergo attack by a water molecule. This is consistent with the result that mutation of U2620(2585) leads to a decrease in the rate of factor-catalysed peptide release²⁴. The requirement for the RF to induce these conformational changes is also consistent with the

similar rate of deacylation stimulated by the RF or tRNA in the presence of 20% ethanol⁶. Furthermore, because unacylated tRNA promotes hydrolysis, despite not having functional groups that could participate directly in the reaction, the primary role of the RF might be to promote the conformational change that we visualize here, and not to activate the water molecule⁶.

Thus, the peptidyl transferase centre may use this induced-fit mechanism to help catalyse both essential reactions in which it participates: the peptidyl transferase reaction during elongation and the programmed deacylation of peptidyl-tRNA at termination. In addition, the structural demonstration that the peptidyl transferase centre uses an induced-fit mechanism to enhance substrate specificity shows that such mechanisms pre-date protein enzymes and were available to the ribozymes of the RNA world.

METHODS

The syntheses of ChPmn, CChPmn and the TSA are as reported elsewhere^{9,25}. Crystals of the large ribosomal subunit of *H. marismortui* were grown and stabilized as described previously²⁶. Crystals were soaked in stabilization solutions containing either 3 mM ChPmn and 1 mM CCApCb, 3 mM CChPmn and 1 mM CCApCb with 100 mM SrCl₂, or 1 mM TSA for between 1.5 and 3.5 h before being flash-frozen in liquid propane. Diffraction data sets were collected at the National Synchrotron Light Source beamline X25 and the Advanced Light Source beamline 8.2.2, using 0.4° oscillations. Data were processed with the program HKL2000 (ref. 27). Structures were determined by rigid-body refinement of the apo-subunit structure (PDB code 1S72; ref. 16) with the program CNS²⁸, followed by iterative rounds of gradient energy minimization, B-factor refinement and modelling with the program O²⁹ (Supplementary Table S1). The set of reflections used for the calculation of R_{free} for these structures is the same as that used for the refinement of the apo subunit²⁶. The complexes were compared (Fig. 3) by the superimposition of phosphorus atoms of domain 5 of 23S rRNA, using the lsq function in the program O.

Received 13 December 2004; accepted 23 August 2005.

- Green, R. & Lorsch, J. R. The path to perdition is paved with protons. *Cell* **110**, 665–668 (2002).
- Kisselev, L. L. & Buckingham, R. H. Translational termination comes of age. *Trends Biochem. Sci.* **25**, 561–566 (2000).
- Koshland, D. E. In *The Enzymes* (eds Boyer, P. D., Lardy, H. & Myrback, K.) 305–346 (Academic, New York, 1959).
- Bennett, W. S. Jr & Steitz, T. A. Glucose-induced conformational change in yeast hexokinase. *Proc. Natl Acad. Sci. USA* **75**, 4848–4852 (1978).
- Caskey, C. T., Beaudet, A. L., Scolnick, E. M. & Rosman, M. Hydrolysis of fMet-tRNA by peptidyl transferase. *Proc. Natl Acad. Sci. USA* **68**, 3163–3167 (1971).
- Zavialov, A. V., Mora, L., Buckingham, R. H. & Ehrenberg, M. Release of peptide promoted by the GGQ motif of class 1 release factors regulates the GTPase activity of RF3. *Mol. Cell* **10**, 789–798 (2002).
- Moazed, D. & Noller, H. F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142–148 (1989).
- Pape, T., Wintermeyer, W. & Rodnina, M. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* **18**, 3800–3807 (1999).
- Katunin, V. I., Muth, G. W., Strobel, S. A., Wintermeyer, W. & Rodnina, M. V. Important contribution to catalysis of peptide bond formation by a single ionizing group within the ribosome. *Mol. Cell* **10**, 339–346 (2002).
- Schmeing, T. M. et al. A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nature Struct. Biol.* **9**, 225–230 (2002).
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
- Agmon, I. et al. Ribosomal crystallography: a flexible nucleotide anchoring tRNA translocation, facilitates peptide-bond formation, chirality discrimination and antibiotics synergism. *FEBS Lett.* **567**, 20–26 (2004).
- Hansen, J. L., Schmeing, T. M., Moore, P. B. & Steitz, T. A. Structural insights into peptide bond formation. *Proc. Natl Acad. Sci. USA* **99**, 11670–11675 (2002).
- Burgi, H. B., Dunitz, J. D. & Shefter, E. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Am. Chem. Soc.* **95**, 5065–5067 (1973).
- Radisky, E. S. & Koshland, D. E. Jr. A clogged gutter mechanism for protease inhibitors. *Proc. Natl Acad. Sci. USA* **99**, 10316–10321 (2002).
- Klein, D. J., Moore, P. B. & Steitz, T. A. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **340**, 141–177 (2004).

17. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA* **98**, 4899–4903 (2001).
18. Weinger, J. S., Parnell, K. M., Dorner, S., Green, R. & Strobel, S. A. Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nature Struct. Mol. Biol.* **11**, 1101–1106 (2004).
19. Schmeing, T. M., Moore, P. B. & Steitz, T. A. Structures of deacylated tRNA mimics bound to the E site of the large ribosomal subunit. *RNA* **9**, 1345–1352 (2003).
20. Seila, A. C., Okuda, K., Nunez, S., Seila, A. F. & Strobel, S. A. Kinetic isotope effect analysis of the ribosomal peptidyl transferase reaction. *Biochemistry* **44**, 4018–4027 (2005).
21. Scarlett, D. J., McCaughan, K. K., Wilson, D. N. & Tate, W. P. Mapping functionally important motifs SPF and GGQ of the decoding release factor RF2 to the *Escherichia coli* ribosome by hydroxyl radical footprinting. Implications for macromolecular mimicry and structural changes in RF2. *J. Biol. Chem.* **278**, 15095–15104 (2003).
22. Rawat, U. B. *et al.* A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature* **421**, 87–90 (2003).
23. Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5**, 1014–1020 (1999).
24. Youngman, E. M., Brunelle, J. L., Kochaniak, A. B. & Green, R. The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell* **117**, 589–599 (2004).
25. Weinger, J. S., Kitchen, D., Scaringe, S. A., Strobel, S. A. & Muth, G. W. Solid phase synthesis and binding affinity of peptidyl transferase transition state mimics containing 2'-OH at P-site position A76. *Nucleic Acids Res.* **32**, 1502–1511 (2004).
26. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
28. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Youngman and R. Green for allowing us to use their unpublished rates of peptidyl-tRNA hydrolysis; G. Muth, D. Kitchen and S. Scaringe for assistance with synthetic chemistry; J. Kavran for collection of a diffraction data set; A. Innis for reading the text; and P. Moore for experimental advice and critical reading of the manuscript. Financial support for the National Synchrotron Light Source comes from the US DOE and the NIH. The Advanced Light Source is supported by the DOE. K.S.H. is supported by a NIH postdoctoral fellowship. This work was supported by NIH and ACS Beginning Investigator grants to S.A.S. and an NIH grant to T.A.S.

Author Contributions T.M.S. determined and evaluated the structures of the three complexes of the 50S ribosomal subunit with substrate analogues under the supervision of T.A.S., and K.S.H. designed and synthesized the substrate analogues used under the supervision of S.A.S.

Author Information The atomic coordinates and structure factors have been deposited into the RCSB Protein Data Bank under accession codes 1VQ6, 1VQ7 and 1VQN. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.A.S. (teather@csb.yale.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Turning the tables

Imagine that China dominated the research world, with the best labs, the most funding and the highest publication record. With little chance for advancement at home, young scientists from the United States and Europe would pour into the country. Because China's research growth has been so rapid, it can't produce enough home-grown scientists. So it welcomes the Westerners — after a fashion.

Western graduate students and postdocs coming into China are expected to teach in Chinese, and if they can't communicate with their Chinese students and colleagues, they lose their funding, their job and their visa.

Westerners would consider such a scenario unacceptable, and with good reason. But that's the situation for many Chinese graduate students and postdocs working and studying in the Western world. The story of Xuemei Han, a Yale graduate student who almost got sent back to China because she was considered "not in good academic standing", despite passing her preliminary exams, is a case in point (see *Nature* **438**, 278–279; 2005). Hundreds of Chinese students rallied to her cause because,

even if they had been successful in the United States, they felt that their presence in the country was tenuous, owing to many circumstances beyond their control. Han eventually managed to secure her grant, found another adviser and kept her visa.

Ideally, Han's predicament will do more than illustrate inequity and will result in better conditions for the foreign scientists much of the Western world depends on. It should rally principal investigators to find ways to make foreign graduate students and postdocs more secure. Solutions could range from more support in language lessons to assurances that visas won't be cancelled if they switch advisers or universities. After all, Western scientists would ask for the same treatment if the tables were turned.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

US Head Office, New York
 345 Park Avenue South, 10th Floor,
 New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
 Peter Bless
 Classified Sales Representatives
 Tel: +1 800 989 7718

**New York/Pennsylvania/
 Latin America:** Andrew Bennie
**Midwest USA/Maryland/
 NIH:** Wade Tucker
East USA/Canada:
 Janine Taormina

San Francisco Office
Classified Sales Representative:
 Michaela Bjorkman
 225 Bush Street, Suite 1453
 San Francisco, CA 94104
 Tel: +1 415 781 3803
 Fax: +1 415 781 3805
 e-mail: m.bjorkman@naturesf.com

European Head Office, London
 The Macmillan Building,
 4 Crinan Street,
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director: Nevin Bayoumi (4978)
European Sales Manager: Andy Douglas (4975)

Advertising Production Manager: Billie Franklin
 To send materials use London address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
 Patrick Phelan
 e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Save the world and keep a career

Throughout his scientific training, Taylor Ricketts was a gung-ho, academic kind of guy. After his postdoc, he faced a fork in his career path, choosing between academia and the 'civil society' world of non-governmental organizations (NGOs). On the promise that he would conduct research, he became director of the conservation science programme at the WWF.

Ricketts, like a growing number of scientists, wants his work to improve the world, and NGOs value research that can help them shape policy. "Research is now considered an important core activity at many NGOs," says Michael Edwards, director of governance and civil society at the Ford Foundation in New York. To get results, he says, NGOs realize that they need to back coherent and convincing stances with research rather than rhetoric.

As the number of NGOs continues to increase — currently some 40,000 international organizations, most of which formed since 1970 — civil society provides an option for scientists, most notably in global health and the environment. As these organizations mature, they are better able to raise budgets that can support research. Oxfam, Save the Children, World Vision and Environmental Defense are just a few NGOs with in-house research departments. Although the number of research positions at NGOs is still low, insiders suggest that the demand for researchers will increase. For those interested in career opportunities at NGOs, it is important to understand their scientific needs, the potential trade-offs and the cultural differences.

Environmental NGOs often have particular, even regional, areas of research interest. For example, Conservation International is based in Washington DC and specializes in rescuing biodiversity. It recently established a tropical ecology, assessment and monitoring project to collect data on climate, soils and biodiversity in South America. The project's scope lured Thomas Lacher from a tenured professorship and endowed chair at Texas A&M University in College Station to become executive director of the organization's Center for Applied Biodiversity Science. More than half of his research staff of 70 have PhDs; the rest have advanced technical backgrounds, showing that scientists are entering the NGO arena at all stages of their careers.

Policy decisions

Most positions requiring a scientific background will translate science into an organization's policy position. Laurie Geller, a science officer at the International Council for Science in Paris, chose to combine her scientific background in atmospheric science with public policy. She helps to create interdisciplinary research programmes and to oversee the production of assessment and advisory reports on scientific issues. She represents the organization at international meetings, for instance, of the United Nations or Organisation for Economic Co-operation and Development (OECD).

Although the number of research positions with environmental NGOs is still small, the need for scientific expertise within the global-health arena is often

If the prospect of endless lab work doesn't appeal, maybe using your qualifications to address global problems more directly would be the answer. There is plenty of scope for those who wish to pursue science with a 'social conscience', as **Virginia Gewin** finds out.



Melissa Pope (above) and Laurie Geller have found work in 'civil society' rewarding.



underestimated. In addition to the need for health professionals, such as epidemiologists and immunologists, at organizations that respond to disasters and emergencies, the global-health NGO community supports a surprising amount of biomedical work to translate basic research into real-world solutions.

Melissa Pope, an immunologist at the Population Council's Center for Biomedical Research, continues work she started at Rockefeller University on HIV transmission to help develop treatments and preventive methods. Her colleagues conduct complementary population biology, demographic and social science studies to enhance treatment effectiveness.

Indeed, several global-health NGOs have found that conducting scientific assessments of the efficacy of various health initiatives helps inform future policies. "I see a growing need for people who bridge boundaries between scientists and policy-makers," says Karin Ringheim, research and analysis director at the Global Health Council in Washington DC. For those wanting to stay in academia, conducting science on policy-relevant areas or volunteering to be a public voice on controversial issues are options (see 'Science policy in academia', opposite). But, for those ready to make the leap from academia to the non-profit world, there are a number of trade-offs to consider.

Money can be an issue. The large established organizations can offer salaries comparable to academic positions, but most smaller organizations,

Taylor Ricketts hopes that his efforts will have an impact on both basic research and on-the-ground conservation efforts.



on tight budgets, offer comparatively low salaries.

In addition maintaining scientific credibility can be a problem. Scientists specializing in environmental or resource-based issues often have their credibility attacked by organizations with opposing views. "Although the aim of NGOs is often overtly political, one should strive to avoid politicizing scientific information and analyses — which can be a real challenge," says Geller.

Time out

Finding time to focus on research is an issue itself. Ricketts, for example, spends only 30–40% of his time doing research. The job-demands inherent to NGOs often outweigh research interests — a common, sometimes frustrating problem. "I don't have time to pursue any basic research. I mostly manage relationships and people," says Xan Augerot, director of science at the Wild Salmon Center in Portland, Oregon.

The biggest consequence of not doing research is limited career mobility. Claudia Neubauer, coordinator of Fondation Sciences Citoyennes in Paris, notes that it is difficult to go back to academia once you have left for advocacy. Ricketts says that this is really his grand experiment — to find out whether he will be able to move back to academia should he choose to do so.

But it can be done. After 20 years at Environmental Defense, Michael Oppenheimer recently became director of the programme in science, technology and

SCIENCE POLICY IN ACADEMIA

There are a number of ways for young scientists to satisfy their idealistic leanings without leaving academia. Sense about Science is a London-based non-profit organization that encourages scientific dialogue on challenging issues. It relies on academics volunteering to take part in public debates or to write documents for lay readers. Recently, young interns have been working as unpaid staff for long stretches of time.

Undergraduates can make a difference. Science Shops — independent or university-affiliated research providers, 70 of them in Europe alone — link non-governmental organizations (NGOs) to scientific expertise (see *Naturejobs* 4–5; 5 July 2001). Supported in part by the European Commission, they provide research free of charge either through volunteer efforts or through credits offered to university students.

"Students want their theses to have a social impact, not just in social sciences, but in economics and natural sciences," says Michael Straehle of the Science Shop in Vienna.

BioVision, a global life-sciences forum, hosts a conference called BioVision.Nxt for young biologists interested in the common ground between academia, industry and NGOs. "We put together the programme to help students and NGOs understand where their mutual interests lie," says Abigail Gemo, director of Biovision.Nxt, in Paris.

environmental policy at Princeton University's Woodrow Wilson School, where he will split his time between policy and science. Oppenheimer says that his ability to continue a research programme at the NGO was key to keeping the door to academia open. But his research — modelling the effects of climate change — was not expensive because it relied on data that had already been generated. "Effective scientists have to know what's going on in the field," he says, "but more importantly they also need to maintain and expand connections with other scientists."

Being a first-rate scientist won't necessarily get you a job. Try volunteering with the NGO to understand the workplace dynamic and demands, or taking a policy internship through a professional association.

Clear thinking

Excellent communication and writing skills, for a variety of audiences, are a must — and brevity is key. Geller is often asked to explain everything about climate change — without jargon — in a five-minute presentation. Seeking opportunities to write for local newspapers or NGOs can be a good first step.

Geller says scientists need to be very adaptable, as they are often asked to jump into topics they know little about. And project management skills will help in juggling the numerous job demands.

Unfortunately, there is no escaping the grant application. "Although the applications from private foundations are often shorter and easier than a typical grant from the National Institutes of Health, the institution gets less money," says Pope.

"This is not a place for someone who needs constant success," says Jane Rissler, food-biotechnology expert at the Union for Concerned Scientists in Washington DC, adding that one must look back at a year-long campaign to see that small victories add up to progress. Much of her satisfaction comes from helping citizens have a role in important policy debates.

Not surprisingly, many people who work for NGOs recommend such work. "When you can make a difference to the lives of many by using your science, you will feel a sense of satisfaction in what you have accomplished that transcends the paper in *Nature* or *Science*," says Gerald Keusch, associate dean for global health at Boston University's School of Public Health. Increasingly, lucky scientists, such as Ricketts, have the opportunity to strive for both.

Virginia Gewin is a freelance writer in Portland, Oregon.

WEB LINKS

Idealist.org job search

♦ www.idealists.org

Sense about Science

♦ www.senseaboutscience.org.uk

Biovision Nxt

♦ www.biovision.org/biovision_16.html

Fondation Sciences

Citoyennes

♦ sciencescitoyennes.org

International Council for

Science

♦ www.icsu.org

Union of Concerned

Scientists

♦ www.ucsusa.org

MOVERS

Richard Somiari, president and chief scientific officer, ITSI-Biosciences, Johnstown, Pennsylvania



2003-04: Chief operating officer, Windber Research Institute, Windber, Pennsylvania

2001-03: Chief operating officer/chief scientific officer, Windber Research Institute/Clinical Breast Care Project

2000-01: Scientific director, Windber Research Institute/Clinical Breast Care Project, Walter Reed Army Medical Center, Windber, Pennsylvania

Richard Somiari's belief that technology can advance biology has taken him from studying food technology in Nigeria to founding a proteomics company in the United States. "Technology has caught up to biology," he says. "I try to adapt technology to answer scientific questions."

Somiari's interest in technology began as a child in Nigeria; he liked to build model cars. His mother wanted him to become a doctor but he was saddened to see the local hospital struggle to cure patients.

"I thought, if I go into research, I can come up with better ways to help patients," says Somiari. He studied microbiology and began working in industry, which at the time in Nigeria was dominated by food manufacturing.

His master's work in enzyme and food technology was his introduction to the study of proteins. Then a fellowship from the Polish government and the World Bank took him to the Technical University of Lodz in Poland and to the University of Strathclyde, Scotland, for a PhD where he switched to biotechnology. He took up molecular biology and biomedical research as a postdoc at the University of Maryland's School of Medicine, studying gene expression in clinical samples.

There he met Nicholas Jacobs, president of Windber Medical Centre in Pennsylvania. Jacobs had funding to establish a research institute and was looking for a scientist to head it. It was an opportunity Somiari couldn't pass up.

"It gave me a chance to finally implement the ideas I'd had over the past ten years," he says. He bought the latest high-throughput genomics and proteomics analytical equipment, established key research partnerships, built an extensive tissue bank and increased the institute's size to 45 employees working on eight research programmes.

By 2004, Somiari saw a need for another organization. Many researchers needed to do analytical proteomics work, but didn't have the necessary instrumentation and couldn't afford to use service companies that cater mainly to drug companies. Somiari started his own company, Integrated Technologies and Services International (ITSI) Biosciences, which began providing affordable proteomics services to academics this year.

Somiari advises young scientists to move into new areas and diversify their knowledge and skills. "The good thing about being a scientist is that they may take your job from you but not your knowledge," says Somiari. "Your success, achievements and failures will remain with you."

Corie Lok

BRICKS & MORTAR

Pittsburgh power tower

The University of Pittsburgh's Biomedical Science Tower 3 (BST3) opened last month. By next year, the \$205-million, ten-storey tower will house a wide range of high-powered nuclear magnetic resonance (NMR) instruments, the world's largest zebrafish colony and a biosafety level-3 lab. There will also be scientists from an eclectic mix of disciplines, from drug discovery to fish genetics and vaccine development. Researchers have worked with architects to make BST3 as user-friendly as possible.

The building will house about 500 workers: some in the structural and computational biology departments; some in neurobiology, bioengineering and molecular genetic groups; and some at centres in neurodegeneration, cognition, drug discovery and vaccine research. Half or so of the 50 principal investigators will be in new jobs.

Angela Gronenborn migrated from the National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland, to chair the new structural biology department and oversee the NMR facility. Her office overlooks the gigantic underground bay holding 600, 700 and 800 MHz magnets (a 900 MHz magnet will be added soon).

The only NMR facility in the United States with such a range of instruments under one roof, it will let researchers elucidate different properties of difficult molecules such as membrane proteins, says Gronenborn. Natural light streams into the bunker, nine metres underground, where removable plate glass windows allow easy installation of future magnets.

Upstairs, the 10,000-tank zebrafish facility can accommodate up to half a million fish, enabling collaborative genetic screening projects. Special light-tight cabinets help speed breeding and experiments. Interchangeable freshwater tanks are equipped with computer-automated pumping and water quality systems.

All that means more time and energy focused on lab work, and less on the 'nuts and bolts' of maintaining the colony, says Nathan Bahary, a zebrafish molecular geneticist at the University of Pittsburgh. The tower also contains modular lab benches and other features for ease of expansion and upgrades.

The diversity of scientists in BST3 — basic 'fish people' collaborating with bioreactor engineers and talking to nearby neurobiologists — should keep the science bubbling.

Kendall Powell

GRADUATE JOURNAL

Who needs evidence?

I have returned to my graduate lab to keep my project moving along and to pass along the scientific torch. Today was a rough one at work and a news report I caught this evening only worsened my mood. It was news from Kansas and it was not pretty. The state's board of education voted to change the science curriculum so that it casts doubt on evolution and includes the teaching of 'intelligent design'. To most scientists, it is a tragedy. To me, it's personal.

I earned my undergraduate degree in biochemistry at the University of Kansas. During those years, I embraced the scientific method with ferocity and moved in the direction of teaching science. Now, this decision in Kansas stands contrary to the principles of science and it feels like a kick in the teeth.

It also inspires some playful mocking and melts away my stress about postdoctoral projects. Let us imagine science free of any burden of evidence. My life would be so much easier! Suddenly, I can come up with sweeping conclusions without doing boring experiments. I am fascinated by fine mechanistic questions in biology and I previously thought such enquiries required rigorous research methods. Now, I realize that biochemical reactions occur because some 'designer' said that's how they should happen.

This could really work in our favour, fellow scientists. Since the news from Kansas, I have finished two 'scientific' manuscripts! I think I can do about ten per week and never have to go to the lab! There's bad news, however. I will scoop all of *Nature's* readership by next Christmas. Wink. Nod.

Jason Underwood completed his PhD in molecular biology at the University of California, Los Angeles, in June.

It never rains in VR

Where two worlds collide.

John Gilbey

Jim and I are a pretty good double act. He's the modeller and I'm the data wrangler — I don't criticize his code and he doesn't mess up my data. Except that we both do, then we argue, then we get drunk, then it's OK. Like last night.

Slouched on the sofa in Theatre 2 this morning, Jim looked even worse than me. This was a lecture theatre back when students actually came here in person. Now it's a virtual reality studio — not the biggest, not the smartest, but it does us very nicely. You see, it saves us having to go out to work.

We got in early on the climate-change ticket, when the University of Rural England was still just a college. Whole-landscape studies, that's our thing. Soil, hydrology, microclimate, plant ecology, you name it. Fifty thousand autonomous data loggers spread across the river basin, all pootling away 24/7 and mesh networked back to the Data Centre.

The really great thing is that it all feeds into the VR system. The whole valley, mapped out in glorious 256-bit colour, photorealistic, surround sound — with the data streamed in real time, or selected highlights, or fast-forward, whatever. So you get to do all your fieldwork from the comfort of a sofa — a real one, you don't mess with things like that — without having to go outside. And it never rains, so everyone's happy.

Except the Prof.

She stamped in just as we got settled. We were in trouble — big time. The model was showing some weird nutrient-flux events that didn't fit the standard profile. Now usually, when things go wrong, Jim and I spot it first and, you know, improve reality a bit. Sadly, our biggest sponsor had seen this one first. The one that pays for all our toys. Oops.

It seems that our data grid now links directly into the government eco-policy engine, so big profile changes have an immediate impact on the national subsidy structure. Nice of her to tell us.

There was the usual 'inside out' feeling as the system came up, then one of the lasers failed to sync. Pulses of purple light don't improve hangovers. Finally, plink, there we were, flying our sofa a thousand

feet above the English countryside. With unnecessary zeal, the Prof zoomed us up the valley to the edge of the moor, dumping us with an inertia-free crash on a convenient rock outcrop. I closed my eyes until the nausea passed.

She spoke to thin air. "Are you there, Jenny?"

A small cartoon rabbit wearing glasses and a T-shirt hopped out from behind a gorse bush. The T-shirt said 'sysadmin' on the front, in pink neon. "Hi folks, want to see the playback?"

Jenny's avatar flipped through some clips showing, on dry summer days, waves of ground water rolling down the hill opposite, like slow liquid avalanches. Glowing numbers and isohyets chased gleefully across the landscape. The system modelled tongues of dissolved nutrients moving at more than two kilometres an hour.

"Can't be happening," I said defiantly, "must be the model."

Jim swore. "Sensor fault — has to be." I pointed out that a dozen devices couldn't all be reading wrong. The Prof steamed. "Just go out there and sort it — now!"

"But it might rain," wailed Jim.

"You won't melt — GO!"

JACEY

So, two hours later, there we were. Looking at the real hill: steep, and smelling strongly of sheep. We stumbled halfway up before Jim's knees went and he collapsed, groaning, on to the grass next to a sensor enclosure. I checked the probe. Automatic diagnostics, self-calibrating, sealed for life: nothing to go wrong. Crazy.

"Afternoon." Looking up, I saw the farmer walking down the hill. He nodded towards Jim. "Not as pretty as the last one." I explained that Jenny was now back in the States, then made some gentle enquiries.

No, he hadn't been spraying slurry. Anyway, all the tractors have loggers, so we'd know, wouldn't we? No, there wasn't a secret underground reservoir and definitely no Roman aqueduct. He paused to adjust his cap.

"Just checking your probe fences," he volunteered, "like it says in the agreement. We start at the top and we work downhill."

We? I glanced uphill and saw his colleague. An old, arthritic sheepdog was gamely stumping towards us, tongue lolling. He was making about two kilometres an hour. Pausing beside the fence he raised one hind leg and, with a look of satisfaction, anointed the sensor with concentrated nutrients.

"Old age," said the farmer. "He has to stop at every one nowadays."

I called Jenny. "Tell the Prof we are witnessing an event. She should be able to see it in the model, sort of." Jenny expressed qualified delight. "She's watching it now — do we need to refine the model?" she asked.

"Only if Vet Science have got some code for a collie with prostate trouble..."

Jenny relayed the information. There were strange rending sounds in the background. "The Prof is biting chunks out of the sofa. I think she's practising for when you get back. Anything else you need me for?"

I pondered this for a moment. "Jenny, when did you last test the fire sprinklers in the VR Theatre?"

John Gilbey is a writer living in the virtual reality of West Wales. The views expressed are his own, and he is welcome to them.

